Visual recognition memory for complex configurations*

ALVIN G. GOLDSTEIN† and JUNE E. CHANCE
University of Missouri, Columbia, Missouri 65201

Previous studies of recognition memory for heterogeneous pictorial stimuli suggest an unusually large storage and retrieval capacity. In the first experiment, three series of homogeneous pictures (faces, ink blots, snow crystals) were presented to six independent groups of Ss, and recognition was tested immediately or 48 h later. Accuracy and sex were related to stimulus configuration; at both time intervals, accuracy was best for faces and poorest for snow crystals. Levels of accuracy were below those attained in studies using heterogeneous arrays. The results of two other experiments suggest the relative unimportance of verbal mediation in recognition of homogeneous pictorial stimuli.

Investigations by Shepard (1967) and Nickerson (1965, 1968), employing hundreds of complex pictures, obtained results suggesting that immediate tests for recognition yield accuracy scores in excess of 90%. Nickerson (1965) points out that, in contrast, the probability of recognizing a second occurrence of a three-digit number, when only six or seven digits intervene between the two presentations, has been found to be roughly 75%. These findings can be interpreted to imply an enormous human capacity to store input when the stimuli are pictorial.

However, it is important to examine an otherwise tacit assumption regarding what these data mean in terms of Ss' ability to recognize. In order to speak logically about "accuracy" or "correctness" in the results of these investigations, in contrast to those of studies of verbal learning, it must be assumed that the correct recognition responses made by these Ss were to the particular, discrete picture, and not to the class membership of the stimulus. In other words, it must be assumed that S reacted to a previously seen stimulus as though it was in totality the identical stimulus shown him earlier. Furthermore, it must be assumed that if the stimulus had been modified—even by a small change within the picture—between the first and second presentations, S could detect that modification and would react by reporting that the stimulus was not the one shown before. Unless this assumption is justified, there is no basis on which to compare these results with the results of investigations using nonpictorial material, because the design of the latter experiments allows E to determine unequivocally if S recognized a particular word or three-digit number

in toto or only in part. If either implicit or explicit comparisons are made between recognition of heterogeneous pictorial and more homogeneous nonpictorial material, then some assumption must be made that S is "totally" accurate when making his correct recognition responses.

We will try to show first the specific reasons why we believe that the studies cited may overestimate the capacity for visual recognition memory. Next, a rationale is presented for designing stimuli more adequate to assess visual recognition performance. An investigation using such stimuli to measure both immediate and delayed recognition memory will be reported. The results of two other experiments into relationships between accurate recognition responses and verbal mediation will be presented. Finally, a discussion of these results will be offered.[1]

Information about the stimuli given in the Nickerson and Shepard reports suggests that the reported accuracy scores could be inflated estimates of recognition memory for pictures. Because of the heterogeneity of items in both the recognition sample and the interference pictures (stimuli were culled from advertisements in magazines), an item could be recognized in the test trial on the basis of its class membership, rather than as a unique stimulus seen on an earlier occasion. In other words, what was stored in recognition memory may not have been details of a particular stimulus but simply a general concept. With heterogeneous stimuli, the concept might be all that was needed for recognition. Moreover, in a heterogeneous series, the effect of the interference pictures may be minor because they do not relate to many pictures in the critical series. Only those pictures sharing common subject matter, i.e., in the same class, will "interfere" with each other.

Obviously, labels for things involve recognition responses; we recognize an automobile in its various shapes, sizes, and colors, and do not normally mistake it for

a bus. When we correctly recognize our own automobile in a crowded parking lot, however, the behavior is no longer the same as class-membership identification (Goldstein, 1958). This kind of recognition is close to that which most people mean when they use the term "recognition." We recognize our child among many children, a face we have seen before among many faces, etc. The point here is that in measuring this kind of recognition memory, stimuli in the experimental series must be highly similar, thus forcing S to make fine discriminations among items with the same class membership but differing only in terms of prior exposure.

Measures of recognition memory for pictures should be as uncontaminated by verbal mediation as conditions permit. Although it may well be impossible to completely stop S from developing verbal codes for visual stimuli, it is possible to reduce the effectiveness of verbal mediation. The investigations under discussion seem especially vulnerable to the criticism that stimulus heterogeneity enhances the usefulness of verbal coding by reducing confusability of verbal labels.

The argument proposed here is quite similar to explanations suggested to account for the fact that recognition memory in general yields higher accuracy scores than does recall. Recognition tasks supposedly measure "partial" as well as complete learning, whereas recall measures only complete learning (McNulty, 1966). The nature of the stimuli in the Shepard and Nickerson studies suggests that partial learning could account for their data.

The method employed here is characterized by extreme homogeneity of stimuli within both the recognition series and the distracting stimuli (see, for example, Rock & Engelstein, 1959). Primarily, we modeled our investigation as closely as possible after studies of recognition of visually presented verbal material. We also wanted to tax the capacity of the recognition system under the limited circumstances of a laboratory experiment. As suggested in the preceding discussion, such stimuli should be homogeneous, i.e., all stimuli within a recognition series from the same psychological class. Interstimulus distances would have to be as small as possible without making any two stimuli perceptually indiscriminable (when viewed simultaneously).

The latter requirement meant that stimuli would have to be fairly complex. Visual complexity is, in part, a function of the number of inversions ("corners" or "turns") in the contours of a configuration (Attneave, 1957b; Goldstein & Andrews, 1962). Another way of viewing the notion
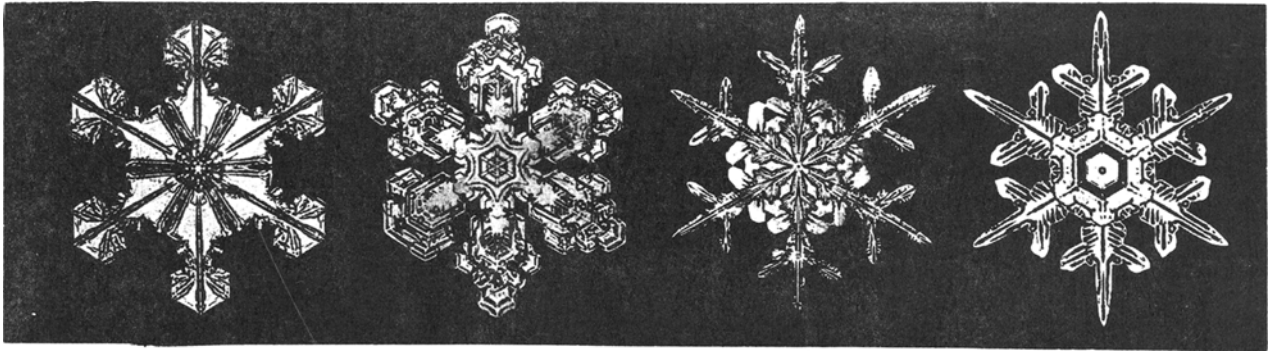
Fig. 1 Examples of the snow crystal photographs used as stimuli.

of complexity is to ask the question: Can an extremely (infinitely?) large number of discriminable stimuli, all still psychologically belonging to the same class, be produced by systematically modifying aspects of the configuration (other than mere size)? If the answer is affirmative, the stimuli are highly complex, because such configurations have enough components that can be varied to produce vast numbers of perceptually discrete stimuli. Finally, the above considerations seem to lead to the conclusion that a stimulus series developed in accordance with these rules would be composed of an infinitely large number of class members. The three configurations actually used in the study—faces, ink blots, and snow crystals—are all naturally occurring configurations which fulfill most, if not all, of the requirements just discussed. In addition, ink blots and snow crystals were specially selected because they, in contrast to faces, should be relatively unfamiliar to our Ss. The face (of the three) is also the visual stimulus most likely stored in memory.[2] Thus, faces were considered to have unique status, and responses to them will be considered to represent recognition memory for a highly overlearned configuration. Comparing the data of the three configurations should give a rough estimate of recognition storage capacity for familiar vs unfamiliar stimuli. Such an estimate must be refined by subsequent research aimed at teasing apart other stimulus factors confounded with familiarity (Attneave, 1957a).

EXPERIMENT 1

Stimuli

All stimuli were black-and-white projected images. Full face pictures of women used in the experiment were selected from university yearbook pictures. Standardization was attained by selecting photographs that differed little, if at all, in terms of head position, camera angle, etc. Noticeable facial blemishes, spectacles, and earrings were all reasons to reject a

photograph. Ink blots were constructed by dripping black India ink upon the crease of a previously folded sheet of white paper (5.5 x 8.5 in.) and then refolding the paper. A blot was rejected as a stimulus only if it had nonstructural defects which could serve as recognition cues, e.g., if during the construction process the blot touched one or more edges of the paper. Enlarged photographs of snow crystals were obtained from a collection of several thousand (Bentley & Humphreys, 1962). In Es' judgments, all stimuli within a class were discriminably different from other stimuli within that class when viewed simultaneously. Pictures within each series were identical in projected size.

Procedure

Each S viewed only one class of stimuli. Procedures were identical for all Ss. In the training session, S was exposed for one trial to 14 stimuli for 2-3 sec, each separated by approximately 5- to 8-sec interstimulus intervals. Order of presentation of the 14 critical stimuli was constant. Prior to the first stimulus, Ss were told that recognition would be tested and that their only task now was to attend to the stimuli. Recognition memory was measured either immediately for half the Ss or after exactly 48 h for the remaining Ss. The method of testing recognition was the same in both groups. In the recognition test, each S was shown for 5 sec each a total of 84 stimuli, of which 70 were "new" and 14 were previously exposed in the training session ("old"). Presentation order of old and new stimuli was random and identical for all Ss across stimulus classes. In this session, each stimulus was identified by a number that was visible above the figure. Responses were recorded by S on a data sheet that included the identifying numbers and two columns headed "Seen Before" and "Not Seen Before." S was never informed of his accuracy. All sessions were conducted in small groups of 5 to 10 Ss. The problem of response bias was handled by explicitly instructing S that there would be 14 "seen-before" stimuli, no more, no less; Ss

were requested to try to restrict the number of their seen-before responses to this number. Compliance with these instructions was facilitated by the open format of the response sheet, clearly numbered for 84 responses, which permitted S to readily count his seen-before responses.

Subjects

Introductory psychology students participated for credit in either one (immediate groups) or two sessions (delayed group). Seventy-eight men and 78 women were assigned in equal numbers to one of six (three stimulus conditions by two delay conditions) experimental groups (N = 26 per group).

RESULTS AND DISCUSSION

Meaningful analysis of responses in a yes-no (old-new) type of test situation is in part dependent upon the so-called false-alarm rate, or the number of old responses made to new stimuli. However, false alarms become of critical importance only if there is no limit on the number of old responses S may emit. If the number of old responses is restricted in some manner, as in this investigation, there is less danger of obtaining spuriously inflated accuracy scores. To the extent that Ss followed instructions, the number of correct old responses should be a fair estimate of recognition. The mean number of old responses, irrespective of correctness, was indistinguishable from the actual number (14) of previously exposed stimuli. Moreover, this mean is very representative, since almost all Ss restricted themselves to 13, 14, or 15 responses.

Analysis of variance (Sex by Stimulus Class by Delay) of the number of stimuli correctly recognized as old by Ss (Table 1) revealed that recognition of the three kinds of stimuli were reliably different (F = 78.62, p < .001). Faces were best recognized (M = 10.0, 71%), followed by ink blots (M = 6.5, 46%) and snow crystals (M = 4.7, 33%).[3] Although the last two

Table 1
Percent Correct Recognition Responses

		Faces	Ink Blots	Snow Crystals
Men	Immediate	66	54	34
	Delayed	65	44	32
Women	Immediate	77	48	39
	Delayed	78	41	28

Note–Values have been rounded to nearest integer.

values appear relatively inconsequential as indices of recognition, further analysis (hypergeometric distribution) indicates that for a *single* S viewing 14 old and 70 new stimuli, correctly guessing as many as 4 stimuli (approximately 28%) is very unlikely (p = .05). Obviously, the probability of two or more Ss correctly recognizing 4 (or more) stimuli by chance is appreciably less than .05. The number of Ss with recognition scores of 4 or greater in each of the six conditions of the experiment is summarized in Table 2, where it can be seen that even in the poorest performing group (snow crystals delayed) most Ss certainly had stored usable information about the stimuli and were able to make reliable recognition responses.

These results suggest that within homogeneous, as opposed to heterogeneous, arrays of stimuli recognition of pictorial stimuli is by no means clearly superior to recognition of nonpictorial stimuli. Moreover, the varying results obtained among the three visual configurations used in this experiment suggest that conclusions regarding recognition of visual stimuli must be dependent, in part, on the particular stimuli employed. In a word, generalizations regarding pictorial recognition may have to be fairly narrow. If one is willing to assume that the face is a highly familiar configuration, when it is placed in a background of similar stimuli the level of performance does not approach the level of performance reported in other studies of visual recognition. Recognition memory for ink blots and snow crystals (assumed to be much less familiar stimuli) can be strikingly poorer.[4] Ignoring, for the moment, all other differences among the three groups of stimuli, subjective familiarity is a major and fundamental discrepancy between faces and the

remaining two classes of pictures. However, ink blots and snow crystals were not equally recognizable, even though both are unfamiliar as visual stimuli and, perhaps, are relatively close together on the familiarity continuum.

As expected, mean correct recognition across all stimuli without delay (7.4) was significantly better (F = 4.13, p < .05) than mean correct with 48 h delay (6.7). The Delay by Stimulus Condition interaction was not significant (F = 1.00). However, overall decrement in performance, which amounted to only 9%, is surprisingly small, considering a single short learning trial and the nature of the stimuli. Moreover, nearly all of this decrement is accounted for by the responses to ink blots ($M_I = 7.1$ vs $M_D = 5.9$ correct) and to snow crystals ($M_I = 5.1$ vs $M_D = 4.2$ correct). Each of these stimuli shows 17% decrement as a function of delay, while responses to faces were unaffected by the 48-h delay (10.0 vs 10.0 correct). These results, in addition to those of other experimental studies of recognition (e.g., Rock & Engelstein, 1959; Clark, 1965; Nickerson, 1968) in which moderate to fairly long delay showed little effect on picture recognition accuracy, suggest problems for further experimentation. Is pictorial recognition memory unusually resistant to performance decrement over time, or is it simply the case that our measurements are too crude to register the effects of delay?

The Sex by Stimulus Class interaction was significant (F = 3.64, p < .05). Women recognize significantly (p < .05) more face stimuli than do men (10.8 vs 9.2), but are no better than men at recognizing the other stimuli (ink blots: men = 6.8, women = 6.2; snow crystals: men = 4.7, women = 4.7). The men's overall score (6.9) was not significantly different from the women's overall score (7.2; F = 1.03, p > .05).[5]

The finding that, with face stimuli, women perform in a manner superior to men is not new. This result has been reported by Witryol and Kaess (1957). It has also been noted by the present investigators in several unpublished studies of face recognition. Note that this difference does not result from a generally better visual recognition on the part of women; the result does strongly imply that

Ss' performance is responsive to differential interest and motivation.

EXPERIMENT 2

The data of Experiment 1 appear to be in agreement with much of the current research and theory on human memory; verbal mediation is critical for memory storage, and the likelihood of verbal mediation increases with increased stimulus familiarity. Although our stimuli were selected specifically to reduce the effectiveness of verbal coding, the basis for the selection was rational and intuitive, not empirical. Experiment 2 was designed to assess verbal responses to the three kinds of configurations used in Experiment 1.

Procedure

The 14 critical (i.e., recognition) stimuli of each of the three classes of stimuli (total = 42) used in Experiment 1 were presented to a new group of 19 college Ss with instructions to give verbal associates to each picture. The instructions were similar to those used in research on associative responses to random shapes (Glaze, 1928; Vanderplas & Garvin, 1959; Goldstein, 1961); Ss were asked to "... write a word or phrase that each picture reminds you of or looks like." If the stimulus had no associative meaning, S was instructed to write "no" in the appropriate space on his data sheet. The actual mode of presentation of the stimuli was identical to that of the first experiment. All Ss saw all three stimulus classes in the order of ink blots, snow crystals, faces.

RESULTS AND DISCUSSION

For the first analysis of these data, only the total frequency (19 Ss responding to 14 stimuli) of response to each kind of stimulus was tallied, except that pure naming responses such as "ink blot," "snowflake," etc., were excluded. Not unexpectedly, responses to faces were most frequent (205), followed by ink blots (179) and snow crystals (132). It was reasonable to conclude, as had others (see Clark, 1968), that ease of recognition is related directly to ease of association; but a more fine-grained analysis, described below, suggests that this conclusion may be misleading.

In Experiment 1, some items within a

Table 2
Number of Ss Per Condition with Four or More Correct Recognition Responses

Faces		Ink Blots		Snow Crystals	
Immediate	Delayed	Immediate	Delayed	Immediate	Delayed
26	26	24	23	22	16

Note–There was a total of 26 Ss in each condition.

Table 3
Relationship (rho) Between Recognizability of a Stimulus and Frequency of Associative Responses

	Faces	Ink Blots	Snow Crystals
Immediate	−.24	+.23	−.13
Delayed	−.18	+.13	−.40

Note–For all coefficients, N = 14, p > .05. All cells are independent. See text for further details.

class in the series of 14 critical stimuli were highly recognizable, whereas other items were not. If verbal mediation were an important factor in recognition, then a relationship should exist between number of associations made to a particular stimulus and number of correct recognitions made to that stimulus. To test this relationship statistically, six rank-order correlations were computed, one for each of the three classes of stimuli within the two levels of delay (see Table 3). Of these six coefficients, only one attained any appreciable size, and it was a negative value (−.40). These relationships do not support a verbal mediation interpretation of differences in recognition memory for the stimuli used in the first experiment. These data are not in harmony with the recent findings of Clark, who reported that association value of random shapes is an important determinant of shape recognition (Clark, 1965, 1968; Clark & Knoll, 1969).

EXPERIMENT 3

Is the instructional set induced in the Ss appropriate to the problem? The association-value measure is probably the measure of meaningfulness most frequently used in memory or learning studies with random shapes (e.g., Price & Hill, 1968; Clark, 1965, 1968; Clark & Knoll, 1969). Perhaps with the present stimuli an association-type response differs from a response made by an observer who is planning to store the information for future retrieval. This possibility was tested in Experiment 3.

Procedure

Two groups of Ss were shown the first 30 stimuli from each of the three classes (total = 90) used in Experiment 1. For one group of 23 Ss, instructions were identical to the "association" instructions of Experiment 2. For a second group of 20 Ss, instructions were ". . . to write something about the picture that would help you recognize it were you to see it again later on. If you cannot think of anything that would help you to remember it, then write the word 'no' in the appropriate space." All stimuli were projected for 2-3 sec; interstimulus intervals were approximately 8 sec. Responses were made in writing on a prepared data sheet. Preliminary findings will be reported here. Results of a more comprehensive study in progress on the same question will be reported in a subsequent paper.

RESULTS AND DISCUSSION

For all stimuli, the number of words in Ss' responses in the "remembering" group is longer than in the "associate" group.

When instructed to "associate," responses tended to be short, unitary, and affective (especially in response to faces). When instructed to "remember," responses tended to be long and conjunctive, with more than one cue for each stimulus. There were fewer "no" responses with the "remembering" set than with the "associative" set. These data imply that in studying verbal mediation to visual stimuli instructional set is a critical variable, among others.

In an attempt to clarify the relationship between the verbal responses to the stimuli and the recognition data, we also looked at the content of the verbal responses made by individual Ss. This analysis indicates that Ss' responses (within a series) to these stimuli appear to be invariant to an extent which makes one question whether mediators of this kind could be helpful in processing and storing pictorial information. Moreover, conclusions from recognition investigations in which the "associative" technique was employed may be misleading to the extent that these kinds of associations are unrelated to the verbal coding elicited by other techniques. Conclusions implying a causal relationship between recognition and verbal coding are now suspect, until there is clear evidence that verbal encoding goes on during the storage process and that the coding is necessary for the recognition performance. Implied here is the possibility that pictures almost always elicit implicit verbal responses from human Ss but that verbal responses are not (or need not be) facilitating for recognition. Considered collectively, these data and the data from many recent investigations of imagery (Paivio, 1969) suggest that pictorial stimuli may not need the assistance of verbal material for storage and retrieval from memory.

Our data, in comparison with data of previous studies, suggest the following: (1) Familiarity with a class of pictorial stimuli improves recognizability of new members of this class and perhaps reduces the effects of "decay" over time. This facilitation may come about as a consequence of frequent input into memory storage and not simply as a consequence of frequently seeing the class members. (2) Familiarity, although a potent factor in recognition memory, might not be the only determinant of recognizability; other stimulus factors play a role in concert with nonstimulus influences, such as, for example, interstimulus similarity of the items within a class, interest of the observer, etc. (3) Some pictorial material may not be facilitated for memory by implicit natural-language verbal mediation. Thus,

the expected increase in verbal coding associated with greater familiarity may not be the key to understanding the concomitant increase in recognizability. Verbal responses to all pictorial material may be ubiquitous, but S may not necessarily make use of the verbal response in the memory process. (4) Attneave's conception of complexity is ideally suited to research with random or generated configurations of all kinds, but it may have serious limitations when applied to naturally occurring configurations. (5) There is need for critical reevaluation of the use of association-value measures in learning experiments, especially in studies of those memorial processes characterized more by discrimination than by generalization.

REFERENCES

ATTNEAVE, F. Transfer of experience with a class-schema to identification of patterns and shapes. Journal of Experimental Psychology, 1957a, 54, 81-88.
ATTNEAVE, F. Physical determinants of the judged complexity of shapes. Journal of Experimental Psychology, 1957b, 53, 221-227.
BENTLEY, W. A., & HUMPHREYS, W. J. Snow crystals. New York: Dover, 1962.
CLARK, H. J. Recognition memory for random shapes as a function of complexity, association value, and delay. Journal of Experimental Psychology, 1965, 69, 590-595.
CLARK, H. J. Random shape recognition at brief exposure durations. Psychonomic Science, 1968, 12, 245-246.
CLARK, H. J., & KNOLL, R. L. Variables underlying the recognition of random shapes. Perception & Psychophysics, 1969, 5, 221-224.
COFER, C. N., & MUSGRAVE, B. S. (Eds.), Verbal behavior and learning: Problems and processes. New York: McGraw-Hill, 1963.
GLAZE, J. A. The association value of nonsense syllables. Journal of Genetic Psychology, 1928, 35, 255-267.
GOLDSTEIN, A. G. On the use of the term "recognition." American Journal of Psychology, 1958, 71, 790.
GOLDSTEIN, A. G. Spatial orientation as a factor in eliciting associative responses to random shapes. Perceptual & Motor Skills, 1961, 12, 15-25.
GOLDSTEIN, A. G., & ANDREWS, J. Perceptual uprightness and complexity of random shapes. American Journal of Psychology, 1962, 75, 667-669.
GOLDSTEIN, A. G., & CHANCE, J. E. Recognition of children's faces. Child Development, 1964, 35, 129-136.
MALPASS, R. S., & KRAVITZ, J. Recognition for faces of own and other race. Journal of Personality & Social Psychology, 1969, 13, 330-334.
McNULTY, J. A. A partial learning model of recognition memory. Canadian Journal of Psychology, 1966, 20, 302-315.
NICKERSON, R. S. Short term memory for complex meaningful visual configurations: A demonstration of capacity. Canadian Journal of Psychology, 1965, 19, 155-160.
NICKERSON, R. S. A note on long-term recognition memory for pictorial material. Psychonomic Science, 1968, 11, 58.

PAIVIO, A. Mental imagery in associative learning and memory. Psychological Review, 1969, 76, 241-263.

PRICE, R. H., & HILL, R. Positive transfer of familiarization training to recognition performance: Effects of stimulus complexity and association value. Psychonomic Science, 1968, 11, 283-284.

ROCK, I., & ENGELSTEIN, P. A study of memory for visual form. American Journal of Psychology, 1959, 72, 221-229.

SHEPARD, R. N. Recognition memory for words, sentences, and pictures. Journal of Verbal Learning & Verbal Behavior, 1967, 6, 156-163.

VANDERPLAS, J. M., & GARVIN, E. A. The association value of random shapes. Journal of Experimental Psychology, 1959, 57, 147-154.

WITRYOL, S. L., & KAESS, W. A. Sex differences in social memory tasks. Journal of Abnormal & Social Psychology, 1957, 54, 343-346.

NOTES

1. Aspects of the subsequent discussion have been mentioned by Nickerson (1965) and Clark (1968) among others; the ideas bear repetition in the present context (see also Chap. 2 in Cofer & Musgrave, 1963).

2. Indeed, the face, in contrast to the other two configurations, is not only a frequent visual stimulus but also one of the more frequently stored configurations. Ink blots and snow crystals, on the other hand, have not only been rare visual stimuli, but individual ink blots or snow crystals most certainly have never been stored in memory by our Ss prior to this investigation. This distinction points to perhaps another dimension of familiarity, viz, some visual configurations are often seen, but individual members are never singled out and stored in memory, whereas with other configurations both of these operations take place.

3. In almost perfect agreement with the face recognition data presented here, Malpass and Kravitz (1969), in a study of recognizability of Negro vs white faces, reported 70% accuracy for white Ss judging white faces and 67% accuracy for Negro Ss responding to Negro faces (all immediate recognition). Their procedures and stimuli (color transparencies) differed from ours, thus increasing the generality of both sets of data.

4. Although it may appear contradictory to the position taken in regard to the Shepard and Nickerson data, ink blots and snow crystals are still impressive demonstrations of capacity. Recognition performance of such high levels in the presence of essentially meaningless stimuli, following a single brief exposure, is intuitively surprising. Perhaps it is not as surprising to find that words are readily stored for recognition. A word is a simple visual structure, a fairly small, unitary input with perhaps three coding systems available into which it may be filed: auditory, visual, and meaning. It can also be rehearsed. In contrast, a picture such as we have used here has many component parts, cannot easily be characterized by a single concept (without the immediate prospect of confusion), cannot be coded auditorily, and the possibility of rehearsal is indeed open to question.

5. With 3- to 14-year-old Ss, no evidence for sex-related recognition accuracy for boys' and girls' faces was found in an earlier investigation (Goldstein & Chance, 1964). The use of faces, therefore, suggests a possible way of investigating the development of sex differences in perception and memory.