ABSTRACT
              Research has demonstrated that subjects are sensitive
to both thematic and non-thematic information in pictorial stimuli.
Three experiments were conducted to investigate memory for pictures
under conditions of difficult foil discriminability and lengthy
retention intervals. The foils differed from the studied persons in
the number and quality of non-thematic details that were added or
deleted. In each experiment, college students viewed colored
photographs, black and white photographs, elaborated line drawings,
and unelaborated line drawings. Subjects (N=81) in Experiment 1 were
given an old/new test either immediately, 1 day, 1 week, or 1 month
after viewing the pictures. Subjects (N=92) in Experiment 2 were
given a 4-alternative forced-choice test at the same time intervals.
Subjects (N=40) in Experiment 3 were given the old/new test 3 months
after viewing the pictures. The results revealed that, with a yes/no
procedure, performance was best on colored photographs, with
performance on each stimulus type additive across the retention
interval. For the forced-choice test, performance on colored
photographs and unelaborated line drawings was best, with performance
again additive across delay. A confusion analysis indicated errors
were based on physical similarity even after 3 months. These results
refute the hypothesis that the memorial representations for pictorial
variations converge to a common, thematic code after lengthy delays;
rather, non-thematic, analogue information is encoded and preserved
for lengthy time periods. (Author/NB)

Title: Long-Term Memory for Pictures under Conditions of

Difficult Foil Discriminability

Authors: Donald Homa and Cynthia Viera

Affiliation: Arizona State University

Mailing Address: Dr. Donald Homa

Department of Psychology

Arizona State University

Tempe, Arizona   85287

Abstract

Memory for pictures as investigated under conditions of
difficult foil discriminability and lengthy retention
intervals. The foils differed from the studied versions in the
number and quality of non-thematic details that were added or
deleted. In each experiment, subjects viewed colored
photographs, black and white photographs, elaborated line
drawings, and unelaborated line drawings, followed by an
old/new (Experiment 1) or 4-alternative forced-choice
(Experiment 2) test given either immediately, 1 day, 1 week, or
1 month later; Experiment 3 replicated Experiment 1 but with a
3 month delay. With a yes/no procedure, performance was best
on colored photographs, with performance on each stimulus type
additive across the retention interval. For the forced-choice
test, performance on colored photographs and unelaborated line
drawings was best, with performance again additive across
delay. A confusion analysis indicated errors were based on
physical similarity, even after 3 months. These results refute
the hypothesis that the memorial representations for pictorial
variations converge to a common, thematic code after lengthy
delays; rather, non-thematic, analogue information is encoded
and preserved for lengthy time periods.

Results from a variety of studies have demonstrated that subjects are sensitive to both thematic and non-thematic information in pictorial stimuli. When permitted to freely inspect a painting, Yarbus (1967) found that subjects fixated 5-6 locations. With additional viewing time, the majority of fixations were not distributed to novel locations but involved refixating the same 5-6 locations, often in the same order. Interestingly, the subject's scan path was markedly altered when the prior set was changed. From these results, Yarbus concluded that the scan path reflected ongoing thought and was the subject's attempt to extract thematic information. In a somewhat different vein, Biederman and his coworkers (Biederman, 1972; Biederman, Glass, & Stacy, 1973; Biederman, Rabinowitz, Glass, & Stacy, 1974) found that jumbling of a photograph differentially affected the subject's ability to identify thematic and object information. More recently, Biederman (1981) has argued that schematic (thematic) information is derived from both objects and the relations among those objects, where the latter may be characterized as involving either physical constraints (e.g., interposition) or referential meaning (e.g., probability). The deleterious effects of jumbling on thematic identification occur because object relations are violated.

What is interesting about picture memory is the potential role thematic and non-thematic information might play in explaining the high levels of accuracy frequently found (e.g.,

Shepard, 1967; Standing, 1973). Advocates of dual-code theory (e.g., Bower, 1972; Paivio, 1971) explain picture memory by proposing that pictures are encoded in two distinct formats, a sensory-spatial-analogue code, as well as in a semantic-thematic-abstract code. The rich visual details of a picture are stored at a sensory level, and the meaning of a picture is stored semantically. Opponents of dual-code theory have argued that subjects do not, in fact, remember the exact visual details and their spatial relations; rather, subjects remember a more abstract representation that captures the meaning of the picture (Anderson, 1985; Pylyshyn, 1984). In fact, Anderson (1978) has demonstrated that dual code and propositional theory can be framed such that their predictions are equivalent; that is, if neither theory is constrained, propositional explanations of picture memory cannot be distinguished from a dual-code interpretation.

Nonetheless, there is surprisingly little empirical data on the long-term retention of pictorial information, at least under conditions of difficult foil discriminability. In an experiment that closely mirrored some of the manipulations of the present study, Nelson, Metzler, and Reed (1974) had subjects first view black and white photographs, elaborated line drawings of these photographs, unelaborated line drawings of these photographs, or appropriate verbal descriptions of these photographs, followed by a 2-alternative forced choice test given immediately and 7 weeks later. On the immediate test, recognition performance was equivalent for the four stimulus types. However, on the test given 7 weeks later,

3   5

performance was significantly worse on the verbal descriptions and equal for the three picture versions. Had subjects stored the visual details available in the picture types, then performance might have been expected to covary with stimulus detail; that is, best on the photographs, next on the elaborated line drawings, and worst on the unelaborated line drawings. Since performance on the three picture versions was equal, and since these versions differed in available visual detial but shared a common theme, support for a nonsensory, propositional explanation seemed clear.

However, there are at least 3 reasons why this interpretation should be questioned: (a) each subject worked with only one of the stimulus types; (b) the foils were semantically unrelated to the studied photographs; and (c) the immediate results were not replicated in a study that employed the same stimuli (Loftus & Bell, 1975). For the purposes of the present study, the first two criticisms are especially important. First, because the degree of discriminability required to distinguish old and new stimuli may not have been equivalent for the different stimulus types, potential performance differences among the three picture formats may have been masked by differences in foil discriminability. In addition, by using only unrelated foils, it is impossible to determine whether performance was mediated by correct retrieval of thematic or non-thematic information. In particular, Nelson et al. could not test a strong prediction of a semantic-only hypothesis: if subjects store only the meaning of a stimulus,

6

can they later discriminate that stimulus from a foil that is theme-preserving but differing in detail? The failure to replicate the immediate results by Loftus and Bell may be due to a number of factors, including the fact that performance was at ceiling (95-100%) in the Nelson et al. study. Unfortunately, Loftus and Bell did not include a delayed test.

Results from recent studies have failed to clarify how pictures are represented in memory. Intraub and Nicklos (1985) found that memory for detailed colored photographs was better if attention was drawn to physical details rather than to meaning details, an outcome opposite to that obtained with verbal materials. Similarly, the picture superiority effect (greater memory for pictures than verbal equivalents) is reduced when a conceptually dissimilar study set is drawn to be schematically similar (Nelson, Reed, & McEvoy, 1977). These studies also employed an immediate test only.

What is needed is an assessment of picture memory over lengthy time periods, where the foils are thematically-similar but differ in detail. Under such conditions, three general predictions can be assessed. If only meaning-preserving information is stored, then performance should be equivalent on picture types that differ only in non-thematic detail, both immediately as well as after lengthy delays (performance should also be quite poor, since originals and theme-preserving foils should be nearly impossible to discriminate). If both thematic and non-thematic information are stored, then performance should be monotonic with the amount and quality of detail contained in the picture. Finally, if both thematic and non-

thematic detail are initially encoded but the non-thematic
detail is rapidly forgotten (or forgotten at a faster rate than
thematic information), then the initial performance advantage
for stimuli high in detail should vanish with sufficiently long
retention intervals; that is, it should be possible to
demonstrate an interaction between pictorial detail and length
of the retention interval.

These predictions were assessed in the present study using
four types of pictorial stimuli; colored photographs, black and
white versions of the colored photographs, elaborated line
drawings of the colored photographs, and unelaborated line
drawings of the colored photographs. At the time of study, the
subject saw different pictures in each of four formats. At the
time of test, the subject had to discriminate studied items
from foils that were thematically similar, differing only in
the number and quality of details from the original. For
example, the subject might initially see an elaborated line
drawing of a stimulus and later be tested on the black and
white version of the same stimulus. To the extent that
thematic detail was held constant in the two versions, correct
discrimination would reflect preservation of non-thematic
detail. These predictions were evaluated both immediately as
well as for delays as long as 1 (Experiment 1 and 2) and 3
(Experiment 3) months.

One other manipulation was used. Recognition performance on
the stimulus variations might vary, depending upon how
recognition was tested. In particular, the processes or

subject strategies evoked by a yes/no and a forced-choice test
might not be equivalent. On a 4-alternative forced-choice test,
the studied item is included among three foils that differ only
in detail. Correct selection of the studied item may involve
comparing each alternative with the available memorial
representation of the stimulus. However, only the available
versions need be considered; that is, the subject need not
consider whether a colored stimulus, different from the one in
the array, was the correct one. In effect, the full set of
potential foils was available to the subject. In contrast, a
yes/no test places a greater (and different) burden on memory.
In order to reject a foil in a yes/no task, the subject must
retrieve information sufficient to reject the stimulus. This
could involve either retrieving details not included in the
presented foil or deleting details from the presented foil.
Importantly, the subject must generate alternatives from memory
rather than evaluating available alternatives (as is the case
in a forced-choice test). For example, if the subject is
presented with a black and white foil of an earlier-presented
colored photograph, correct rejection may hinge upon retrieval
of the appropriate colored details. This potential difference
in strategy was tested by using both kinds of tests following
an identical study session; in experiments 1 and 3, a yes/no
procedure was used, whereas a 4-alternative forced-choice test
was used in experiment 2. Potential differences in strategy
for these two types of test are discussed in more detail in the
General Discussion.

7

Experiment 1

Method

Subjects. A total of 81 introductory psychology students
served as subjects to complete the course requirements. Twenty
subjects were randomly assigned to one of four different delay
groups; immediate, 1 day, 7 days, and 28 days. One subject in
the immediate delay group was replaced as an outlier because
her scores were 3.94 standard deviations below the group mean.
Subjects were tested in groups of five.

Stimuli. The stimuli consisted of 96 color photographs
depicting a salient theme and non-essential detail, and three
sets of variations. For each colored photograph, three
versions were prepared, a black and white copy, a detailed line
drawing, and a sparsely detailed line drawing. The detailed
line drawing was generated by tracing over the color photograph
and included the non-essential detail but not the shading or
fine-grained detail. The undetailed line drawing was generated
by tracing over the detailed line drawing so as to preserve the
salient theme of the original color photograph but eliminating
all non-essential detail. The four sets of 96 items were made
into 35 mm slides and presented on a Kodak carousel projector.
An example of the four versions is shown in Figure 1.

----
insert Figure 1 about here

Items were counterbalanced across subjects using a 4 X 4
Latin square, so that each item was presented for study in each

of the four formats in all four delays. Four different

presentation orders were randomly generated so that five

subjects at each delay received the same presentation order

during study. Additionally, four different presetnation orders

were randomly generated and used in a similar manner for the

test presentations.

Design and procedure. A 4 (delay) X 4 (stimulus type) mixed

design was used, with the delay factor as a between-subject

variable and stimulus type as a within-subject variable.

During study, all subjects were shown all 96 themes, 24 in

each of the four formats, for 5 sec. Subjects were seated in a

semi-circle approximately 6 feet from a white projection screen

suspended from the wall. Those subjects in the immeidate

condition were given an intervening arithmetic task for

approximately 12-15 min immediately following the study session

while the test slides were being arranged. The testing

procedure for the four delays was otherwise identical.

At test, subjects were again shown all 96 items, 50% in the

same format as during study, and the remainder as foils,

equally distributed among the three remaining formats. Thus,

of the 24 items originally viewed as color slides, 12 appeared

again as color slides during the test session. Of the

remaining 12, four appeared as black and white slides, four as

detailed line drawings, and four as sparsely detailed line

drawings. This procedure was replicated for all four formats.

Subjects were informed prior to the test that 50% of the

stimuli would appear in the exact format as during study and

that the remaining 50% were equally likely to appear in any of

the other three formats. Subjects were told to answer "old"
for only those items that appeared in exactly the same format
as during study, and to respond "new" otherwise. Pre-test
instructions included admonitions that theymust answer all
trials, guessing when necessary. Further, subjects were
required to give a confidence rating for each answer using a
scale of 1-3, where 1 = uncertain, 2 = moderately certain, and
3 = very certain. The test items were presented at the rate of
5 sec per slide.

Results

Performance measures based on hit rates, a d' analysis
(Green & Swets, 1966), and an accuracy measure combining
correct recognitions and correct rejections (with and without
the confidence judgments) all resulted in identical statistical
outcomes. Analyses based on d' are reported.

Figure 2 shows the mean d' score for each stimulus type as a
function of the four delays.

----
insert Figure 2 about here

In general, performance decreased across the month delay by 50-
70%, with overall d' scores highest for the color photograph
(2.08), relative to the other stimulus types (black and white =
1.42; unembellished = 1.34; embellished = 1.07). The main
effects of delay and stimulus type were both significant, $F(3,
76) = 24.00$ and $F(3, 228) = 32.46$, respectively, each $p < .001$.
The interaction between delay and stimulus type was not
significant, $p > .10$. Post-hoc Tukey tests revealed that

performance on the color format exceeded performance on the remaining stimulus types (p < .01), with both black and white and unembellished better than embellished (p < .05). A simple analysis on the immediate test only showed that the superiority of the color format was evident on this test, p < .001.

The decrease in performance across delay was characterized by comparable changes in both hits and false alarms; overall, hits decreased by .150 (immediate = .778; 1 month = .628), and false alarms increased by .153 (immediate = .157; 1 month = .310). Even after a month delay, performance on each stimulus type was better than chance (p < .01 in each case).

Confusion Analysis. A confusion analysis was done to determine if the tendency to false alarm was a function of the study/test format, and whether these rates changed over time. For example, 24 colored photographs were presented in the study phase, and 12 appeared as old items in the test phase. The other 12 studied photographs appeared equally often as foils in the other three formats (BW = black and white; E = embellished; U = unembellished). Table 1 shows the false alarm rates for each study and test format at each delay, conditionalized upon an error; the chance confusion rate is .33.

—————————————————————
insert Table 1 about here
—————————————————————

A number of interesting results are apparent. First, confusions are not random but are related to the similarity of the study format. For example, on the immediate test, colored

photographs and black and white photographs are frequently
confused, as are elaborated and unelaborated line drawings.
Second, when colored photographs and unelaborated line drawings
were studied, the best foil type remained the same across the
four delays (for colored photographs, the best foil throughout
was the black and white photographs; for the unelaborated line
drawings, the best foil throughout was the elaborated line
drawings). In contrast, the best foil for black and white
photographs and elaborated drawings changed over time. For
example, the best foil on an immediate test for the black and
white photographs was a colored photograph, but after one
month, the best foil was an elaborated line drawing. Finally,
confusion rates are disproportionate for each stimulus type
even after one month, except for the elaborated line drawing
stimuli.

Discussion

Recognition accuracy was significantly better for colored
photographs, relative to the other formats, regardless of time
of test. Furthermore, there was no evidence of an interaction
between stimulus type and delay, which suggests that the codes
representing each stimulus type were stable; that is, initial
representations did not converge into a common format, at least
not after a month's delay. Further support for the stability
of each format across delay was found in the confusion
analyses: even after one month, confusion rates were non-
uniform across format.

In Experiment 2, a 4-alternative forced-choice test replaced
the yes/no task used in Experiment 1. This task should

maximize discriminability difficulty, since the correct choice
was always paired with its three variations. To the extent
that irrelevant detail is rapidly forgotten over time, then
performance should rapidly decline to a common level at near-
chance levels. Should stimulus type again be above chance at
each delay, and should performance on stimulus type be additive
with delay of test, then evidence for a stable representation
based on distinguishing visual detail would again be found.

Experiment 2

Method

Subjects. A different group of 92 introductory psychology
students served as subjects in Experiment 2. A total of 12
subjects were replaced as follows: In the immediate delay
group, one subject was replaced for his failure to answer all
trials and one subject for his failure to maintain the correct
numerical response sequence. In the 1-day delay group, three
subjects in one testing session were replaced for their
inability to follow instructions; additionally, one subject
failed to answer all trials, another failed to maintain the
correct numerical response sequence, and an entire group of
five was replaced due to the interruption of the testing
session by one of the subjects who was ill and wished to
discontinue the session. Subject assignment was identical to
Experriment 1.

Stimuli. The same four stimulus sets from Experiment was
used in the same manner during the study session. A fifth set
of slides consisting of 95 of the original 96 themes was

prepared in the form of a four-item array containing the four
different stimuli versions (the absence of one theme from the
test set was due to the inadevertent failure to photograph the
test array for that theme). For all 95 test slides, the array
format was as follows: the color version was shown in the upper
left quadrant, the black and white version was shown in the
upper right quadrant, the detailed line drawing in the lower
left quadrant, and the sparsely detailed line drawing in the
lower right quadrant.

   Design and Procedure. The design and study procedure were
identical to Experiment 1. At test, subjects were shown all 95
test arrays and were instructed to identiy the stimulus they
had viewed during the study session. They were informed that
during the study session, approximately 25% of the stimuli had
appeared in each of the four formats, and that, for each test
array, a color photograph would appear in the upper left
quadrant, a black and white photograph would appear in the
upper left quadrant, a detailed line drawing in the lower left
quadrant, and a sparsely detailed line drawing in the lower
right quadrant.

   Subjects were instructed to carefully inspect each test
array and to indicate which of the four stimuli had been viewed
at study by recording their choice on a prepared scoresheet.
They were then told to indicate their confidence, for their
first choice only, by using the 1-3 confidence scale described
in Experiment 1. After deciding their first choice and
assessing their confidence, subject were then told to rank-

order the remaining formats from most likely to least likely.
If subjects were unable to rank-order their remaining choices,
they could ignore them. However, they were instructed that for
each trial they must always indicate their first choice and
their confidence rating for that choice.   .

Subjects were told they would have about 10 seconds on each
trial, and that, prior to the end of the trial, they would hear
the word "respond". If they had not made a response by that
time, they were then to write down their choices, guessing if
necessary. Care was taken to insure that all subjects
understood the instructions prior to commencing the test
session. The test items were presented at the rate of about
11-12 seconds per item. Periodically during the test session,
subjects were given the current trial number to insure that
their responses were being notated in correct numerical
sequence.

Results

Performance measures based on hit rates, a d' analysis, and
an accuracy measure combining correct recognition and correct
rejections all resulted in identical statistical outcomes.
Analyses based on d' are reported.

Figure 3 shows the mean d' score for each stimulus type as a
function of the four delays. In general, performance decreased

----

insert Figure 3 about here

----

across the month delay by about 60%, with overall d' scores
highest on the color photographs (1.56) and unembellished

drawings (1.56), relative to the other two stimulus types
(black and white = 1.04; embellished = 1.32). The main effects
of delay and stimulus type were both significant, $F(3, 76) =$
21.35 and $F(3, 228) = 35.91$, respectively, each $p < .001$. The
interaction between delay and stimulus type was not
significant, $p > .10$. Post-hoc tests revealed that performance
on the color format and unbellished format did not differ, $p >$
.10, but both exceeded performance on the remaining stimulus
types ($p < .01$), with embellished better than unembellished, p
$< .01$.

The hit rate decreased by about 30% for each stimulus type
across the month delay (immediate = .716; 1 month = .430).
Nonetheless, performance on each stimulus type was better than
chance even after 1 month ($p < .01$). For example, poorest
performance was obtained on the black and white photographs
after one month, yet all 20 subjects provided a positive d'
value for this stimulus type after 1 month.

Confusion Analysis. Table 2 shows the confusion rate for
each study and test format, as a function of delay

_____

insert Table 2 about here
_____

(conditionalized upon an error). As was the case in Experiment
1, confusion rates appear to be systematically related to the
original study format, even after 1 month. The confusion rate
for the best foil does show some decline across the retention
interval (immediate = .504; 1 day = .539; 1 week = .460; 1
month = .432), but these rates are still above chance, even for

the 1 month condition, $p < .01$.

Discussion

The results of Experiment 2 were similar to those obtained in Experiment 1, with one major exception: performance on colored photographs was again best, but the ordering on the other stimulus types failed to match that obtained in Experiment 1. In particular, performance on the unelaborated drawings was now equal to that of the colored photographs, and performance on the black and white photographs was the poorest. Otherwise, no evidence for convergence among the four types across the delays was obtained, which again argues against a convergence to a common code for the stimulus types. This conclusion is supported by the confusion data, in that confusions are not uniformly distributed among the formats.

Experiment 3

Experiment 3 attempted to replicate the results of Experiment 1 (especially the ordering of stimulus types), and investigated the potential for convergence among stimulus codes for a far greater delay (3 months rather than 1 month). In addition, a within-subject manipulation of delay was used, with different sets of stimuli used on the immediate and delayed test.

Method

Subjects. A total of 40 introductory psychology students served as subjects, none of whom had served in the previous two experiments.

Stimuli. The same stimulus assignment used in Experiment 1

17

was used in Experiment 3, with half of the stimulus set (48
items) used in the immediate test and the remaining 48 used in
the 3 month delayed condition. The four stimulus types were
counterbalanced across subjects over the two delays, using a 4
X 4 Latin square design similar to Experiment 1. Eight
different presentation orders were randomly generated and used
during the study sessions and an additional set of eight
different presentation orders were randomly generated and used
during the test session.

Design and procedure. A 2 (delay) X 4 (stimulus type)
within-subject design was used. Study and test procedures were
identical to Experiment 1, with the exception that no
confidence ratings were collected in Experiment 3.

Results

Performance measures based on hit rates, d', and recognition
accuracy all resulted in identical statistical outcomes.
Analyses based on d' are reported.

Figure 4 shows the mean d' score for each stimulus type as a
function of delay. In general, performance decreased by about
80% at the 3 month test, with overall d' scores highest on the
color photographs (2.02), relative to the other stimulus types
(black and white = 1.33; embellished = 1.01; unembellished =
1.12). The main effects of delay and stimulus type were both
significant, $F(1, 78) = 146.57$ and $F(3, 156) = 16.98$,
respectively, each $p < .001$. The interaction between delay and
stimulus type was not significant, $p > .10$. Post hoc tests
revealed that performance on the color format exceeded

performance on the remaining stimulus types (p < .01), with performance on the remaining formats not differing from each other (p > .10).

Hit rates decreased by about 30% across the 3 month delay (immediate = .780; 3 month = .492), and false alarms increased by about 16% over the same period (immediate = .184; 3 month = .348). Although the performance drop was considerable across the 3 month delay, performance was still above chance, with 34 of the 40 subjects having positive d' values (p < .01). Of the four stimulus types, only the elaborated drawings functioned at a chance level after 3 months (d' = 0.10, p > .10).

Confusion Analysis. Table 3 shows the rates of false alarms for each stimulus type as a function of the format originally studied, shown separately for immediate and delayed tests.

----
insert Table 3 about here

As was the case in the previous experiments, false alarms on the immediate test are strongly related to the visual similarity between original and test format; on the test 3 months later, false alarm rates are a bit erratic but, overall, are still disproportionately related to visual similarity. The major difference on the delayed test is the reduced false alarms to black and white photographs when the studied item was a colored photograph, and the reduced rates to unembellished drawings when the original was an embellished drawing. However, colored photographs are false alarmed at high rates to studied black and white photographs, and unembellished drawings

are frequently called old when embellished stimuli were studied. Overall, it would appear that the rate of false alarming to visually similar items is reduced after 3 months, relative to an immediate test, but that the rate is still greater than chance.

Discussion

The results of Experiment 3 largely replicated those found in Experiment 1. Therefore, the performance ordering on a yes/no test appears to be reliable one, and one that is unchanged across delayes as long as 3 months after study. Furthermore, no evidence for a convergence to a common code was apparent after 3 months.

General Discussion

Three major results emerged from the present study: (a) subjects encode sufficient pictorial detail in a single 5 second viewing period to reject thematically-related foils over time periods as long as 3 months later; (b) the memorial representation of thematically-related stimuli does not converge to a common code after lengthy delays; and (c) the type of test strongly influences performance on stimulus types.

In each experiment, performance on the four stimulus types was above chance, with the sole exception of the elaborated line drawings in Experiment 3 at the 3 month delay. These results are all the more remarkable, given that subjects observed each of 96 stimuli for only 5 seconds and were later required to discriminate the studied stimuli from theme-

preserving foils after substantial delays. At a minimum,
subjects clearly store more than core or thematic information;
some degree of irrelevant and/or idiosyncratic detail is also
encoded and retained. The additivity of stimulus format and
retention interval suggests that information does not decay
into a common, thematic code, with non-thematic detail decaying
at a faster rate than thematic information. Rather, the
pictorial code decays uniformly for all types of detail. An
alternative hypothesis is that the non-thematic detail decays
at a faster rate, but this decay rate is compensated for by the
encoding of a vast amount of non-thematic detail. Further
support for the durability of distinctive codes comes from the
kinds of false alarms made by subjects. Even after 3 months,
errors occur disproportionately to visually similar foils.
There is a suggestion that the rate of visually-based
confusions is slowly declining with time; only large-scale
parametric tests (varying study time and using longer delays
such as 6 months) would verify whether this trend is specious
or real. Regardless, the stability of confusions and the
obtained additivity of stimulus type with retention interval
seem at variance with studies using verbal materials; for
example, the convergence of performance on verbatim and gist
recall after brief delays (less than 1 hour) has been obtained
frequently in studies of sentence memory (e.g., Reder, 1982).
These studies typically use short sentences with little detail,
in which subjects are asked to discriminate tense change and/or
changes in which foils contain the subjects or verbs from other

studied sentences. We are currently investigating memory for short passages that have substantial embellishment of detail. Under these conditions, retention for detail might better mimic that found for pictures.

It is tempting to conclude that these results favor a dual-code interpretation to a propositional one. However, Anderson's (1978) argument is a compelling rejoinder to such an interpretation. Nonetheless, these results place severe constraints on the kind of propositional theory needed to explain these findings. A propositional theory which states that only core information is either initially stored or available after a short duration (e.g., 1 hour) is clearly incorrect. Anderson (1985) has contended that: "...subjects are not likely to remember the exact visual details or spatial relations in a picture. Instead, they are remembering some rather abstraction representation that captures the picture's meaning" (pg. 107). The results of the present study suggest that this contention is misleading and, perhaps, incorrect. Rather, the propositonal code must be sufficiently complex to accommodate both thematic and non-thematic detail for time periods as long as 1-3 months after brief inspection.

The results obtained with the two paradigms (yes/no and forced-choice) did not produce the same ordering for the various stimulus types. This outcome, although unexpected, appears to be stable, since the ordering obtained in Experiment 1 was replicated in Experiment 3. In Experiments 1 and 3, a yes/no procedure resulted in an ordering that was generally consistent with the number of details available at study:

colored photographs were recognized more accurately than black and white photographs, which enjoyed a slight (non-significant) advantage over the line drawings (both elaborated and unelaborated).   In Experiment 2, a 4-alternative forced-choice test favored colored photographs and unelaborated line drawings, followed by elaborated drawings and black and white photographs.   Since the experiments were otherwise identical, an explanation based on differences in strategy or processing seems likely.

In a yes/no task (Experiments 1 and 3), the subject must evaluate the match between the current stimulus and its memorial represenation.   Except for cases when the subject either failed to encode the study item and/or failed to retrieve its representation at test, the match between a test probe and its representation must always be positive, since all foils were variations on studied stimuli.   To evaluate the goodness of the current match, the subject must generate, from memory, variations of the current stimulus probe by adding or deleting details.   If any of the generated matches exceed the match of the current probe, then the subject is likely to respond that the stimulus is "new".   Importantly, the generated matches are likely to exceed the match of the current stimulus only if the subject adds/deletes the correct detail.   For example, the subject presented with an elaborated line drawing at the time of test might first match this stimulus with its memorial trace.   If the original (study) item was a colored photograph of the same theme, successful rejection of the

current probe requires that the subject retrieve the appropriate coloration of the stimulus. The successful addition of detail is required whenever the test probe is less complex than its studied variation. When the test probe is more complex than its studied variation, correct rejection requires the proper deletion of detail. Thus, a test probe that appears as a colored photograph can be rejected as "new" only if the subject correctly deletes the proper detail (e.g., removing color if the original probe was a black and white photograph). A number of subjects wrote comments reflecting the addition/deletion of details, e.g., "I remember that the dress was a deep blue" (addition of detail). Since accuracy in the yes/no procedure favored performance on the more detailed stimuli, and sicne this advantage was uniformly maintained across a month's delay, it must be concluded that the additional detail inherent in the more complex stimuli was encoded and faithfully preserved over time.

In contrast, the forced-choice test (Experiment 2) precludes the generation, from memory, of alternatives, since the choices available to the subject are prsented in the test array. On a forced-choice test, the subject may compute similarity matches for each test variation, identifying as "old" the one version receiving the highest score. What needs explanation is the ordering of performance of the variations, especially in light of the different ordering on the yes/no task. One possibility is that the colored photographs and unelaborated line drawings, which occupy extreme positions on an assumed scale of complexity, were subject to less interference/competition than

the two intermediate variations. For example, colored
photographs differ from black and white photographs only in the
absence of color; from elaborated line drawings in terms of
color and precision of detail; and from unelaborated line
drawings in terms of color, precision of detail, and number of
details. In effect, the colored photographs contained features
that differed from the foils in one, two, or three ways (color,
precision, number). A similar argument may be made for the
unelaborated line drawings; its foils also vary in one, two, or
three ways. However, both the black and white photographs and
the elaborated line drawings have two foils that differ in only
one way. For example, an elaborated line drawing differs from
black and white photographs in terms of precision of detail,
and from unelaborated line drawings in terms of number of
details. The existence of two highly-similar foils for the
elaborated line drawings and the black and white drawings,
versus the existence of one highly-similar foil for the colored
photographs and unelaborated line drawings, may be the basis
for the performance ordering. Note that this rationale, which
predicts the obtained ordering on the forced-choice test
(Experiment 2) is inappropriate as an explanation for the
yes/no task, since, in the latter, the addition/deletion of
details must be generated from memory. If the full complement
of alternatives is not generated, and/or the accuracy of
retrieved details is imperfect, then the number of close foils
in the yes/no task may be functionally smaller than in the
forced-choice task.

Although neither paradigm produced evidence of convergence
to a common, thematic code after 1-3 months, it may be argued
that the distinction between thematic and non-thematic detail
is not as sharp as supposed. This argument might suggest that
the full complement of visual details are not incidental
attributes but part of the intrinsic core meaning of a
stimulus. Thus, coloration may produce moods that modify the
core meaning or theme of a stimulus, as might the very
angularity/curvilinearity/texture of details. A propositional
code, that is sensitive to all details and their remarkable
preservation over time, could accommodate the present results.
This interpretation, however, is inconsistent with both the
scanning literature and its emphasis on thematic encoding
(e.g., Yarbus, 1967), as well as any assertion that only core
meaning is maintained (e.g., Anderson, 1985). At a minimum,
these results provide an empirical foundation which any theory
of pictorial encoding must explain: after 5 sec of inspection,
most subjects (85%) can discriminate theme-preserving
variations from each other 3 months later. What is needed is
an analytical technique which captures the full complement of
detail encoded by the subject. Asking subjects to report
detail may vastly underestimate what has been stored, since
subjects may be unable or unwillilng to report the visual
nuances available to them. At a more theoretical level, the
results of the present study are not necessarily inconsistent
with a propositional explanation; however, these results place
severe constraints on the form such a theory must take. Core
information is clearly not the sole product of encoding, and

the resulting memory trace clearly preserves fine visual detail over lengthy delays. The adequacy of any theory of visual representation, including propositional accounts, must ultimately satisfy a generation test: the information supposedly extracted and maintained in memory must lend itself to a generation process that produces a visual copy that the subject cannot dis riminate from recollection of the studied version. The present results suggest that the development of such a model would be a formidable task, since the number and quality of preserved details and their relations are vast.

References

Anderson, J. R. (1978). Arguments concerning representations for mental imagery. _Psychological Review,_ 85, 249-277.

Anderson, J. R. (1985). _Cognitive psychology and its implications._ New York: W. H. Freeman.

Biederman, I. (1972). Perceiving real-world scenes. _Science,_ 177, 77-80.

Biederman, I. (1981). On the semantics of a glance at a scene. In K. Kubovy and J. R. Pomerantz (Eds.), _Perceptual organization._ Hillsdale, NJ: Lawrence Erlbaum Associates.

Biederman, I., Glass, A. L., & Stacy, Jr., E. W. (1973). Searching for aspects in real-world scenes. _Journal of Experimental Psychology,_ 97, 22-27.

Biederman, I.,Glass, A., Rabinowitz, J. C., & Stacy, Jr., E. W. (1974). Searching for objects in real-world scenes. _Journal of Experimental Psychology,_ 97, 22-27.

Bower, G. H. (1972). Mental imagery and associative learning. In L. Gregg (Ed.), _Cognition in learning and memory._ New York: Wiley.

Green, D. M., & Swets, J. A. (1966). _Signal detection theory and psychophysics._ New York: Wiley, 1966.

Intraub, H. & Nicklos, S. (1985). Levels of processing and picture memory: The physical superiority effect. _Journal of Experimental Psychology: Learning, Memory, and Cognition,_ 11, 284-298.

Loftus, G. R. & Bell, S. M. (1975). Two types of information in picture memory. _Journal of Experimental Psychology:_

Human Learning and Memory, 104, 103-113.

Nelson, D. L., Reed, V. L., & McEvoy, C. L. (1977). Learning
to order pictures and words: A model of sensory and semantic
encoding. Journal of Experimental Psychology: Human
Learning and Memory, 3, 485-497.

Nelson, T. O., Metzler, J. & Reed, D. A. (1974). Role of
details in the long-term recognition of pictures and verbal
descriptions. Journal of Experimental Psychology, 102, 194-
196.

Paivio, A. (1971). Imagery and verbal processes. New York:
Holt, Rinehart, & Winston.

Pylyshyn, Z. W. (1984). Computation and cognition. Cambridge:
MIT Press.

Reder, L. M. (1982). Plausibility judgment versus fact
retrieval: Alternative strategies for sentence verification.
Psychological Review, 89, 250-280.

Shepard, R. N. (1967). Recognition memory for words,
sentences, and pictures. Journal of Verbal Learning and
Verbal Behavior, 6, 155-163.

Standing, L. (1973). Learning 10,000 pictures. Quarterly
Journal of Experimental Psychology, 25, 207-222.

Yarbus, A. L. (1967). Eye movements and vision. B. Haigh,
(trans.). New York: Holt, Rinehart and Winston.

Author Identification Notes

Footnotes

1. This is an implicit assumption of most forced-choice tests, and one that was made explicit in the present study. In fact, one could construct a forced-choice test that included some, but not all, alternatives, and presented foils differed from the original in both type (e.g., black and white) and distortion of type (e.g., a colored foil of an originally studied colored stimulus). Manipulation of the latter type of foil would provide a fine-grained test of the amount and quality of preserved detail.

2. Having subjects rank order the remaining alternatives after they had made their first selection resulted in many blank or undecipherable responses for the rank order data. For that reason, that data are not reported here. The confidence data for the first selection did not suffer from these scoring problems, but the analysis of confidence data produced statistical outcomes identical to that reported in the results section.

Table 1

False Recognition Confusions, Conditional Upon an Error

Test Stimulus

	Imm				1 Day				1 Week				1 Month		
P	BW	E	U	P	BW	E	U	P	BW	E	U	P	BW	E	U
---	.434	.247	.319	---	.492	.297	.261	---	.471	.279	.250	---	.462	.231	.308
.525	---	.250	.225	.361	---	.347	.292	.354	---	.369	.277	.284	---	.500	.216
.156	.221	---	.624	.250	.280	---	.470	.217	.317	---	.467	.254	.373	---	.373
.118	.176	.706	---	.246	.261	.493	---	.235	.235	.529	---	.183	.338	.479	---

4

35

Table 2

False Recognition Confusions, Conditionﾭl Upon an Error, Experiment 2

Test Stimulus

		Imm				1 Day				1 Week				1 Mo	
		P	BW	E	U	P	BW	E	U	P	BW	E	U	P	BW	E
	P	---	.487	.313	.201	---	.530	.232	.238	---	.556	.204	.240	---	.430	.321
Study	BW	.500	---	.300	.200	.520	---	.259	.239	.372	---	.376	.252	.430	---	.300
	E	.103	.369	---	.528	.193	.284	---	.572	.205	.336	---	.459	.212	.342	---
	U	.152	.345	.502	---	.104	.361	.535	---	.154	.392	.454	---	.225	.389	.386

Table 3

False Recognition Confusions, Conditional Upon an Error,

Experiment 3

Test Stimulus

| | Immediate | | | | 3 Months | | | |
	P	BW	E	U	P	BW	E	U
P	----	.742	.194	.065	----	.302	.245	.453
BW	.367	----	.429	.204	.462	----	.280	.258
E	.229	.208	----	.562	.196	.370	----	.435
U	.255	.191	.553	----	.449	.202	.348	----

38

Figure Captions

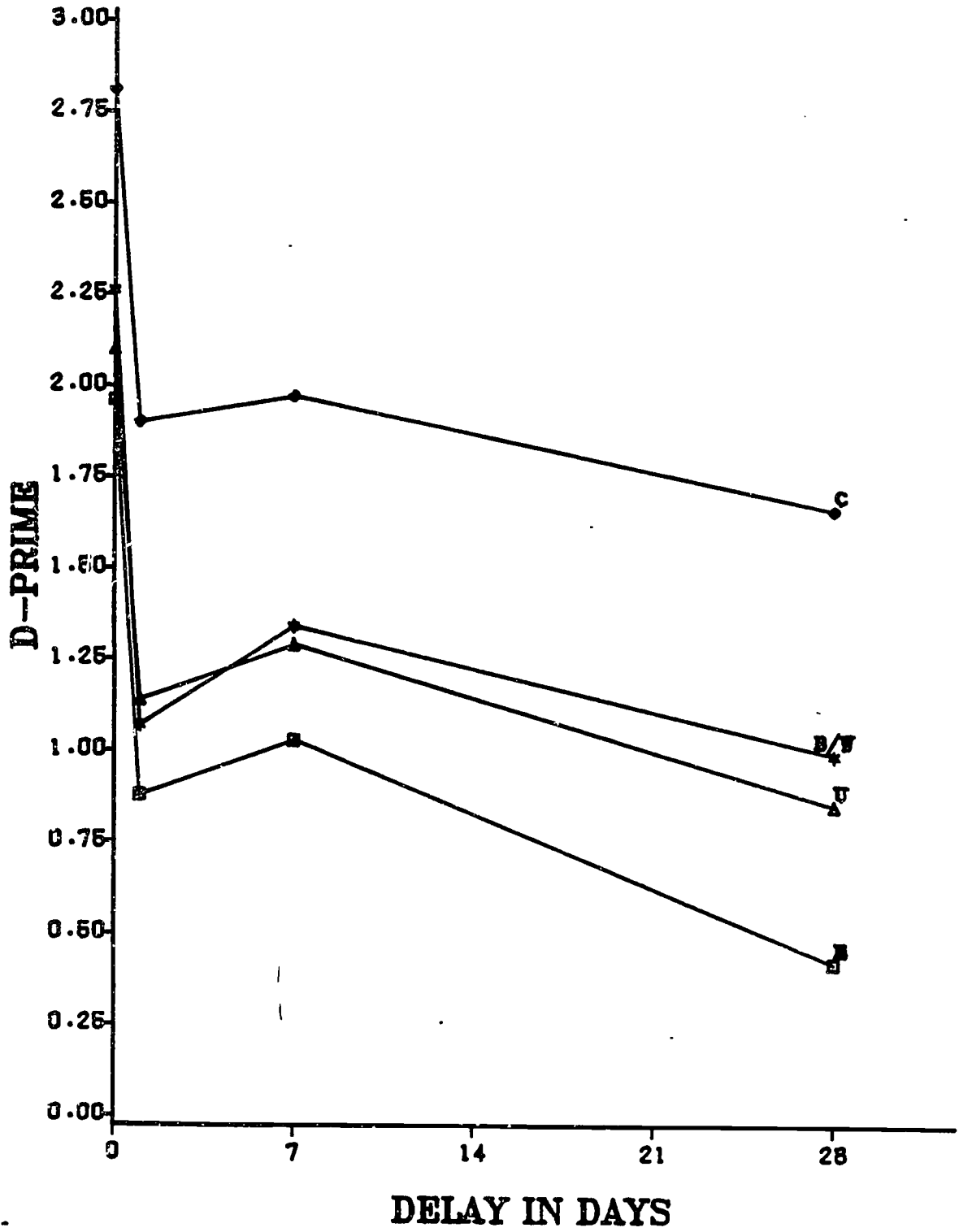Figure 1. An example of the four stimulus versions

Figure 2. Mean d' for each stimulus type, as a function of retention interval (C = colored photographs; BW = black and white photographs; E = elaborated line drawings; U = unelaborated line drawings), Experiment 1.

Figure 3. Mean d' for each stimulus type, as a function of retention interval (C = colored photographs; BW = black and white photographs; E = elaborated line drawings; U = unelaborated line drawings), Experiment 2.

Figure 4. Mean d' for each stimulus type, as a function of retention interval (C = colored photographs; BW = black and white photographs; E = elaborated line drawings; U = unelaborated line drawings), Experiment 3.

40

SINGLES D-PRIME

41

ARRAYS D-PRIME

d'

4.0

3.5

3.0

2.5

2.0

1.5

1.0

0.5

0.0

C

B/W

U
E

0 7 14 21 28 35 42 49 56 63 70 77 84

DELAY IN DAYS

43