Human Memory Capacity for Object and Scene Representation

Aude Oliva
Department of Brain and Cognitive Sciences
Massachusetts Institute of Technology
oliva@mit.edu
http://cvcl.mit.edu

In collaboration with
Timothy Brady, Talia Konkle & George Alvarez

Research funded by NSF CAREER IIS # 0546262
What’s the Capacity of Human Memory?

What we know...

Standing (1973)
10,000 images
83% Recognition

... people can remember thousands of images

What we don’t know...

... what people are remembering for each item?

According to Standing

“Basically, my recollection is that we just separated the pictures into distinct thematic categories: e.g. cars, animals, single-person, 2-people, plants, etc.) Only a few slides were selected which fell into each category, and they were visually distinct.”
What’s the Capacity of Human Memory?

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Fidelity</th>
</tr>
</thead>
</table>
Massive Memory I: Thousands of objects
Massive Memory Experiment I

A stream of objects will be presented on the screen for ~ 1 second each.

Your primary task:

**Remember them ALL!**

*afterwards you will be tested with...*

---

**Completely different objects...**

**Different exemplars of the same kind of object...**
Massive Memory Experiment I

Your other task:

Detect exact repeats anywhere in the stream
Massive Memory I: Methods

Showed 20 observers 2560 unique objects from 480 different object categories.

Number of objects per category varied from 1 to 16.

N-back, detect exact repeats, 2 to 1024 back.

Followed by 240 2-alternative forced choice tests.
Bagel
Backpack
Bucket
Camera
Cup saucer
Tent
Watergun
Recognition Memory Performance

Prediction 1
Catastrophic Interference

Percent Correct

93 ± 1
89 ± 1

Number of studied exemplars within each category

Novel Exemplar

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>4</th>
<th>8</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>Novel</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exemplar</td>
<td></td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
</tr>
</tbody>
</table>
Recognition Memory Performance

Prediction 2
Strong Interference

Percent Correct

93
89

Number of studied exemplars within each category

Novel Exemplar

Chance level

1 2 4 8 16

93 89
Recognition Memory Performance

Prediction 3
No Interference

Percent Correct

<table>
<thead>
<tr>
<th>Number of studied exemplars within each category</th>
<th>Novel</th>
<th>Exemplar</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>93±</td>
<td>89±</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

"Number of studied exemplars within each category"
Recognition Memory Performance

Prediction 4
Minimal Interference

<table>
<thead>
<tr>
<th>Number of studied exemplars within each category</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>Novel</td>
</tr>
</tbody>
</table>

- 93 ± 89
Recognition Memory Performance

Highly Detailed
Minor Interference

Percent Correct

93
89
88
86
82
82

Number of studied exemplars within each category

Novel Exemplar
Recognition Memory Performance

Highly Detailed
Minor Interference

82
16
What about detection performances?

- During exposure phase, N-back repeat detection task probes recognition performances like an old/new task (e.g. familiarity).

- Have you seen that exact same image before in the stream?
Repetition Detection Performances

N-back Performance by # of items back

% correct (HIT)

95 94 87 79 63

# intervening items

% False Alarms

... high performance is not just about the 2-AFC memory test.
What about object similarity?
Category Interference Measure

Interference Score = slope of line

Interference Score = slope of line

Percent Correct

No interference
Min interference
High interference

Number of subsequent items within category

Novel | Exemplar

0 | 1 | 3 | 7 | 15

0 | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100

Percent Correct
Category Interference Measure

Computed Separately for Each Category

Percent Correct

Number of subsequent items within category

No interference
Min interference
High interference

Novel | Exemplar

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
<th>3</th>
<th>7</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>90</td>
<td>90</td>
<td>90</td>
<td>90</td>
<td>90</td>
</tr>
</tbody>
</table>
Does **Distinctiveness** in the category make it easier to remember more items?
I – Measure of **Conceptual** Distinctiveness

Are There Few or Many **Kinds**?

Each category was judged by 12 observers on a 1-5 scale.
I – Measure of **Conceptual** Distinctiveness

**Similar**

**Distinctive**
II – Measure of **Shape** Distinctiveness

How Similar or Different are their **shapes**?
II – Measure of **Shape** Distinctiveness

**Similar**

**Distinctive**
III – Measure of Color Distinctiveness

How Similar or Different are their colors?
III – Measure of Color Distinctiveness

Similar

Distinctive
Distinctiveness vs. Interference

Conceptual Distinctiveness helps you remember

Conceptual

Interference Slope

Bin Number

Similar

Distinctive

$t(196)=2.18, p<0.05$
Distinctiveness vs. Interference

**Color**

No Effect of Perceptual Distinctiveness

- \( t < 1, p = 0.92 \)

**Shape**

- \( t < 1, p = 0.64 \)

No Effect of Perceptual Distinctiveness
Why have one Massive Memory Experiment,

when you can have two?

Contact (1997)
how far can we push the fidelity of visual LTM representation?

*Same object, different states*
Massive Memory II: 2500 unique object categories

1-back

1024-back

Followed by 300 2-alternative forced choice tests

100 novel pairs
100 exemplar pairs
100 state pairs
Examples of Exemplars in memory test
Examples of *states* in the memory test
Results Memory Test

No interference!

92 %

87 %

87 %
How many different images can you see before losing familiarity?

Power law $(r^2 = .988)$

for $n = 1024$ (79 %)

Other function: $r^2 = .988$

Percent Correct

Number of Intervening Items ($n$)

false alarm rate 1%
What about distinct textures?

d′ ~ 0
Concluding Remarks – Part I

Capacity of Human Representation

- Can be massive and detailed
- details are not by necessity discarded through visual transformations

Structure of Human Memory

- Memory for “visual” details is linked more to conceptual knowledge rather than perceptual similarity
How detailed are visual scene representations?
At a glance ... You remember the category and the layout but you have lost some object details

You have seen these pictures

You were tested with these pictures [average false alarms ~ 30%]
What about memory for thousands of scenes?

128 unique semantic categories of natural images

Presentation: 3 seconds each
Beach
Bedroom
Cavern
Closet
Countryroad
Greenhouse
Hair salon
Iceberg
Library
Waves
Methods – The Study Stream

128 unique semantic categories of natural images

2912 natural images shown in the stream (3 seconds each, 800 msec ISI)

**Number of exemplars** per category: 4, 16, or 64!

N= 24 observers
Online Task: **Detect Exact Repetates**

Repeats could be 2 to 1024 back in the stream
Methods – The Memory Test

Followed by 224 **2-alternative forced choice tests**

**Novel** | **Exemplar**
---|---

Test Pairs were always the same for all subjects (4 test pairs for each scene category)

Any effect of interference is due to the additional exemplars seen in the stream
Results – Memory Test Performance

Exemplar

Replication of Object experiment

Percent Correct

0 10 20 30 40 50 60 70 80 90 100

1-novel 4 16 64

96 ± 84 ± 80 ±
Results

Highly Detailed
Minor Interference

Percent Correct

100
90
80
70
60
50
40
30
20
10
0

1-novel
4
16
64

1-novel

4

16

64

Exemplar
Highly Detailed
Minor Interference

Percent Correct

Exemplar
Massive Memory for Scenes and Objects

Memory Performance Comparison

- Scene
- Object

- Chance

Percent Correct

~3000 exemplars

Number of Exemplars (log scale)
Distinctiveness among exemplars

High homogeneity

High Heterogeneity

63 %

91 %
Standing out details... the novelty factor

Accuracy: 100 %
Accuracy: 70 %
Confusion from the Mean ...

Accuracy: 62 %

Very typical images are confused
What is this massive visual memory capacity good for?
Models of Object Recognition

- A massive memory for details lend credence to object recognition approaches that require brute force storage of *multiples* viewpoints and exemplars (and image alignment approaches)
Recognizing the gist of a scene

Proposal: Massive memory capacity is the infrastructure of scene gist recognition.

The brain perceives ~ 60 millions diagnostic inputs per year (3 samples per second).

A robust representation of natural images require accumulated information about the details.

... the challenge for natural image recognition systems is to find the relevant regularities to encode.
Human Scene Understanding: What are the rules of memory distortion?

*Photographic Memory* = A unique code per image

Image perceived

**Memory Distortion**
Compression - Reconstruction

"ocean"  
Basis of spatial layout properties

Conclusion

Memory capacity for natural images is of an order of magnitude higher than previously believed.

Fidelity of storage of visual details is very high.

A unique “conceptual hook” permit to store images with preserved featural details.