VSS 2007
Symposium on Scene Perception

Minimal Scenes, Maximal Challenges

Irving Biederman

University of Southern California

Collaborators:
Ken Hayworth
Xiaomin Yue
Ed Vessel
Rosemary Behizadeh
Robert Mezzanotte
Gary Klatsky

NSF 04-20794, 05-31177, 06-17699

Some Issues of (Novel) Scene Representation

1. **What defines a scene (vs. a collection of objects)?**

   Plausible answer: Relations.

2. **Proposed taxonomy of relations:**

   A. Spatial proximity (“and” “next to”) [May be effective only with a very small number (two or three) of objects.]

   B. “Prepositional” (“in”, “on”) or “gerundal” (-ing verbs).

   C. Familiar

3. **Challenge: Scene Interpretation (not just setting classification).**

Memory for interacting pairs (lower set) much better than non-interacting images (upper set).

(William) Epstein, Rock, & Zuckerman (1960).

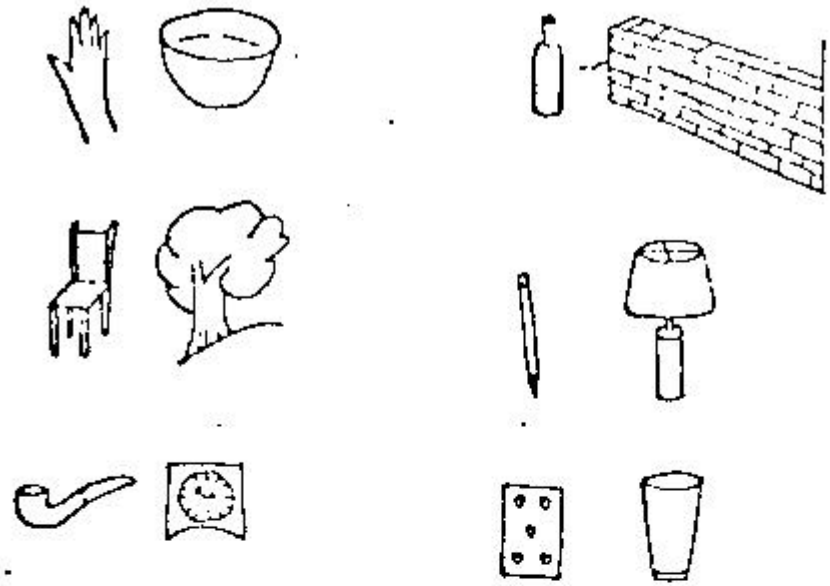


FIG. 3. Pairs learned by Group 2, Experiment VIIb.

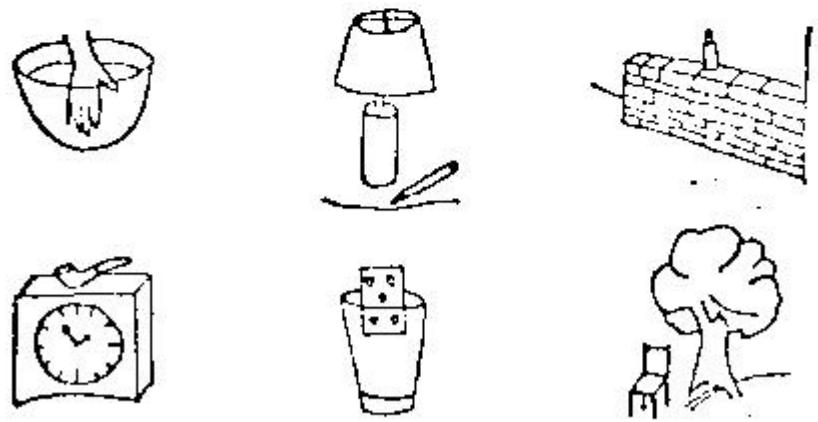
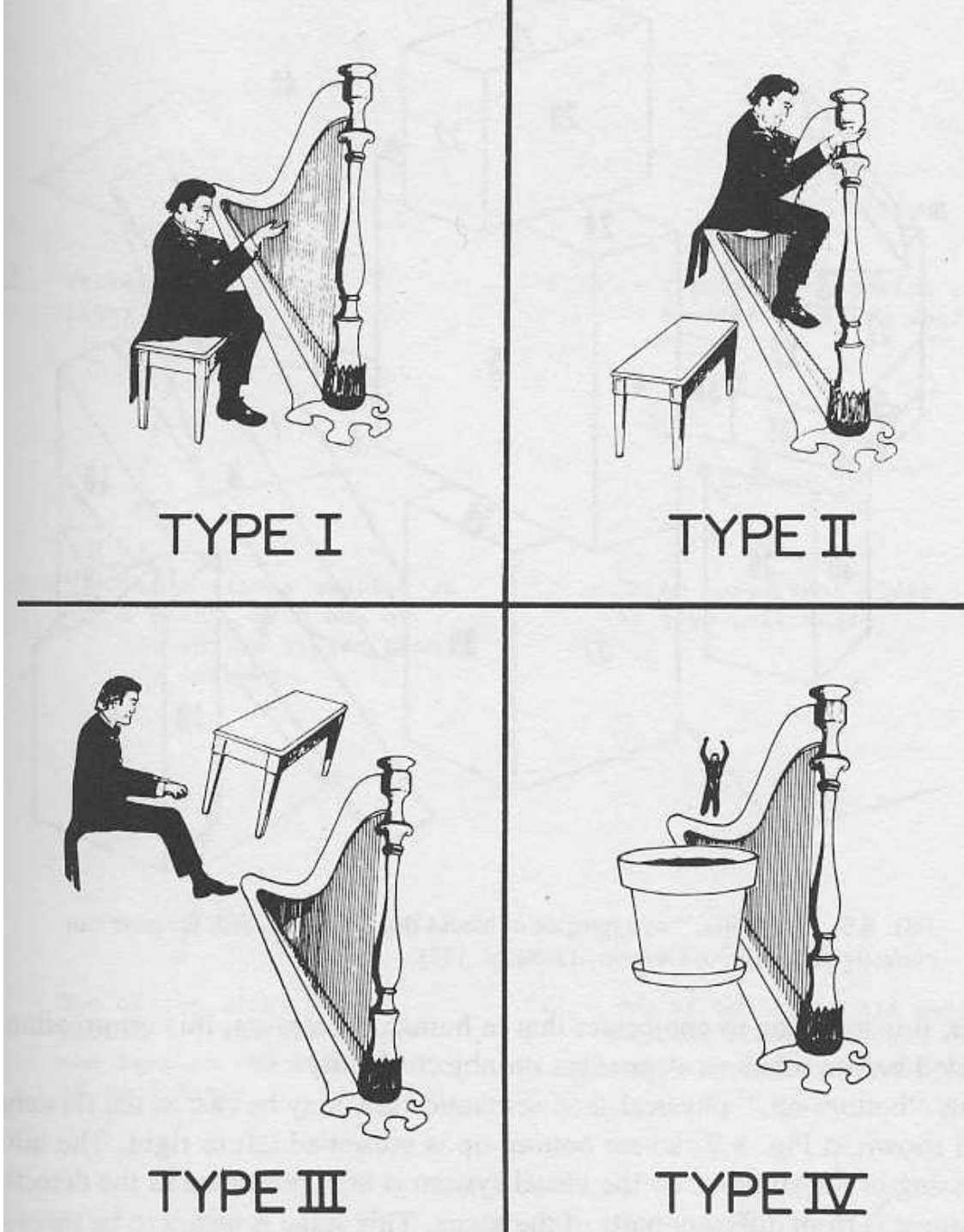


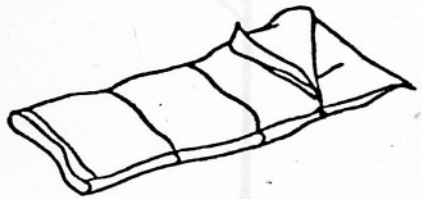
FIG. 4. Pairs learned by Group 1, Experiment VIIb.

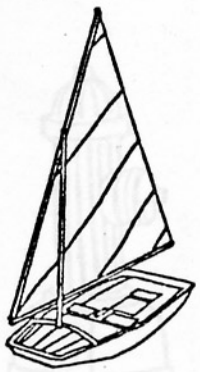
As more “violations” of familiar scene relations--of position, size, support, probability, and interposition--are incurred, the display more and more starts to resemble a collection (of unrelated objects) rather than that of a well-formed scene.

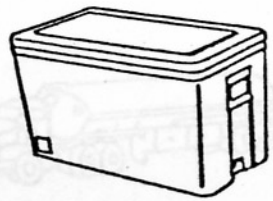
From Hoch et al. (1977)

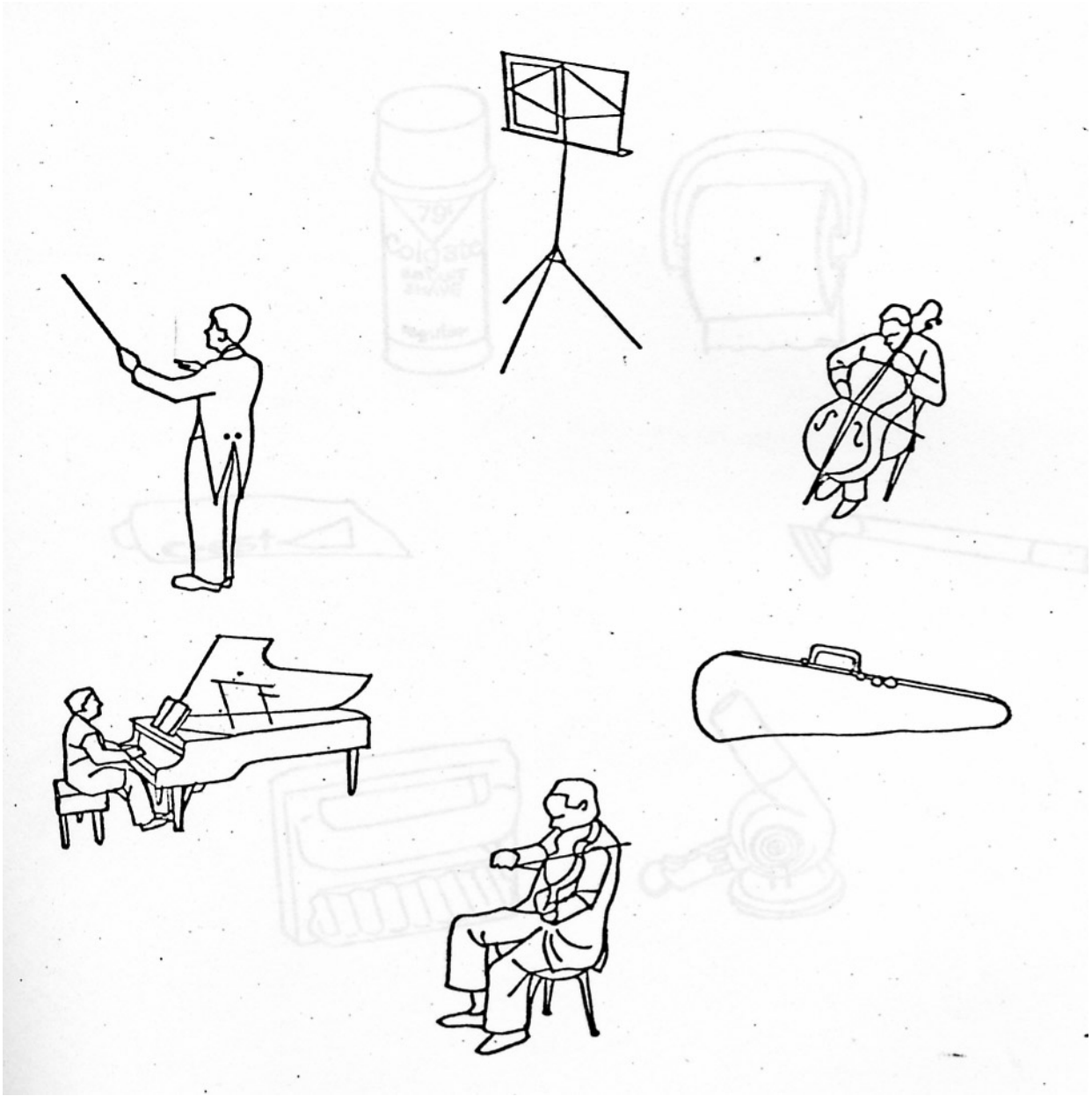


Is the cost of object recognition in a collection of objects a function of the number of objects in the array?

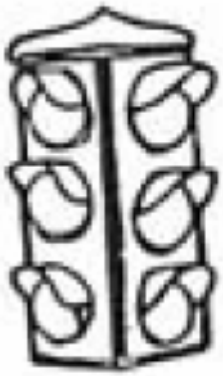


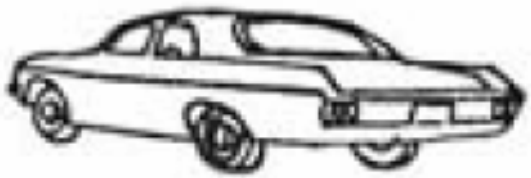


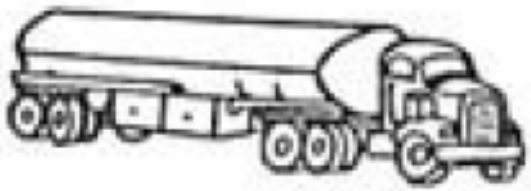




Mailbox



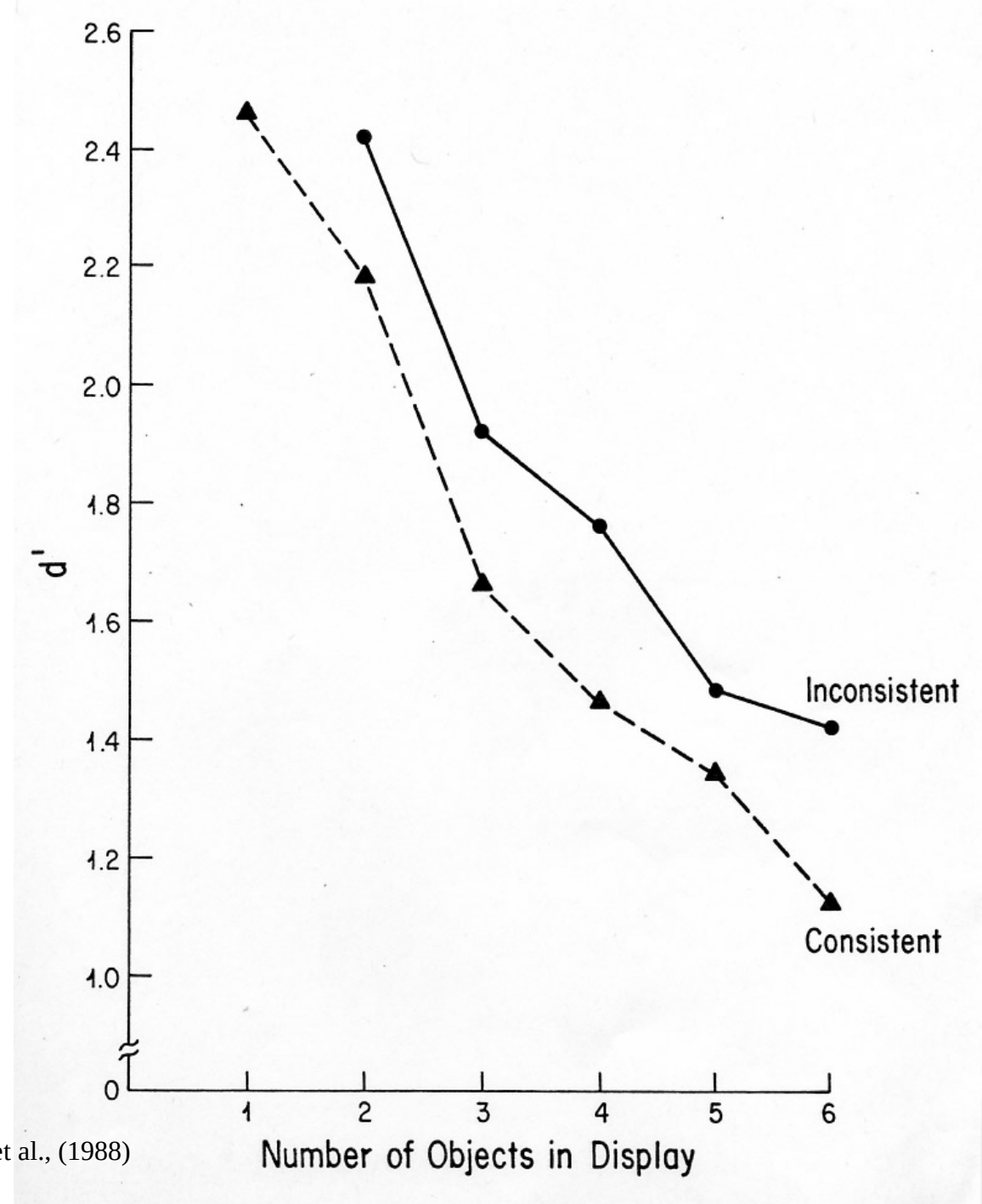




Filing Cabinet

Performance accuracy declines monotonically with each additional item in the display (100 msec exposure).

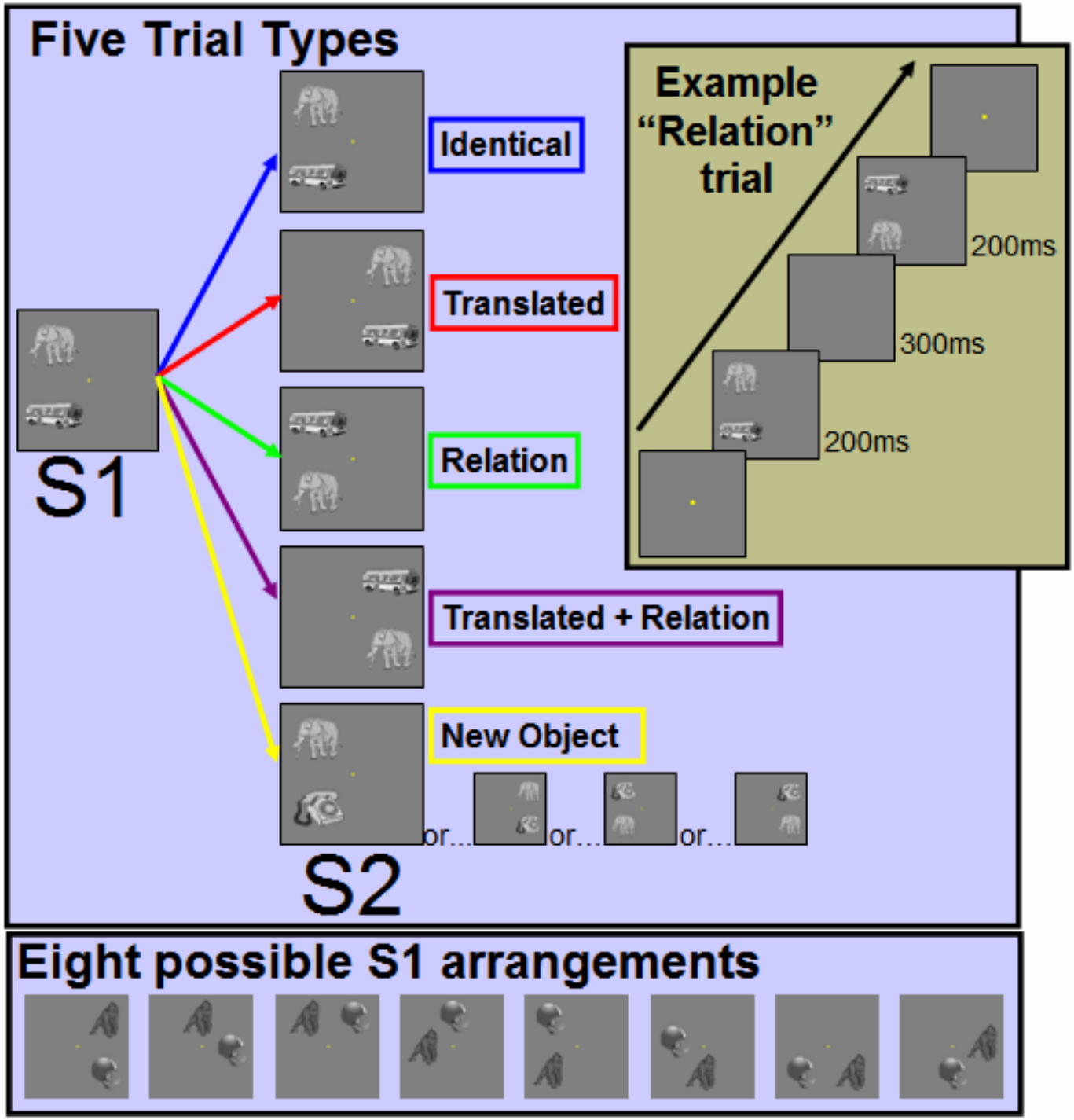
It is possible that when objects are in a prepositional, gerundal, or familiar relationship, the costs of object number are reduced.

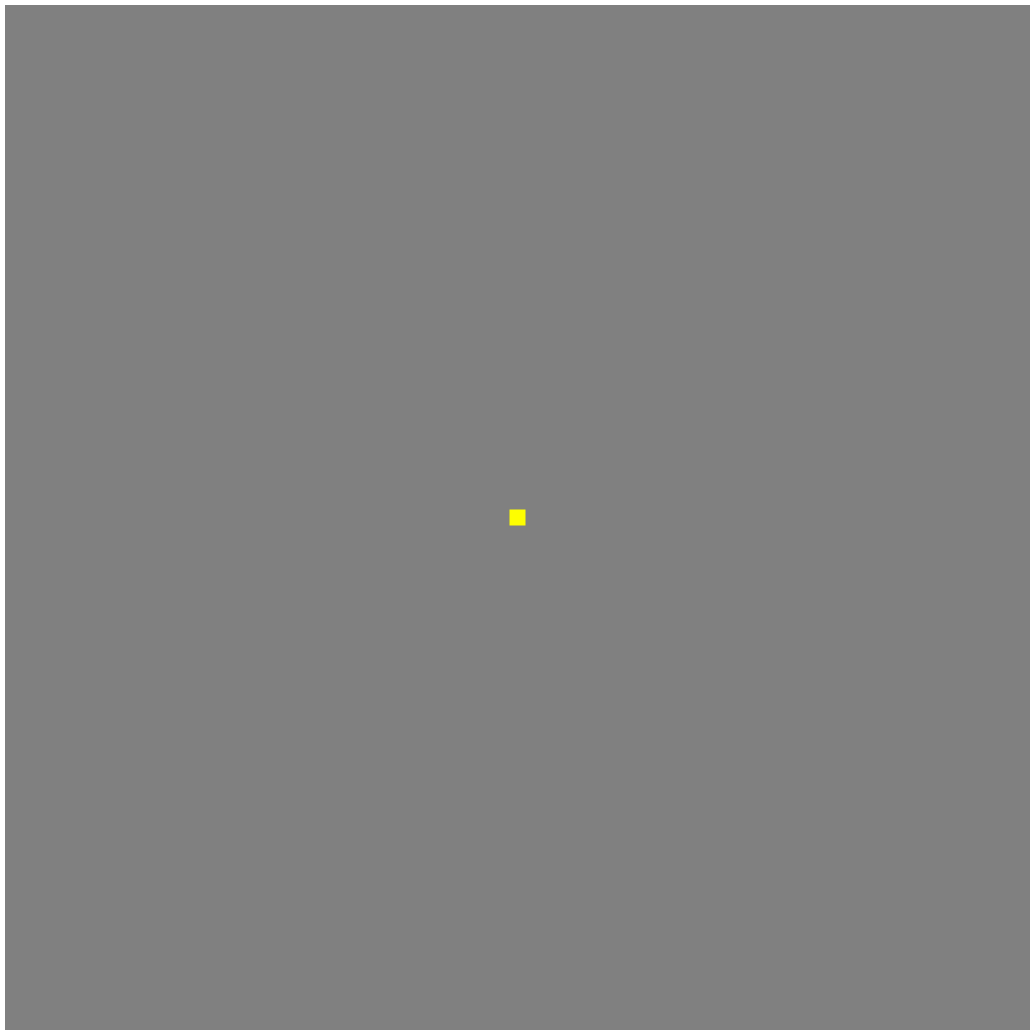


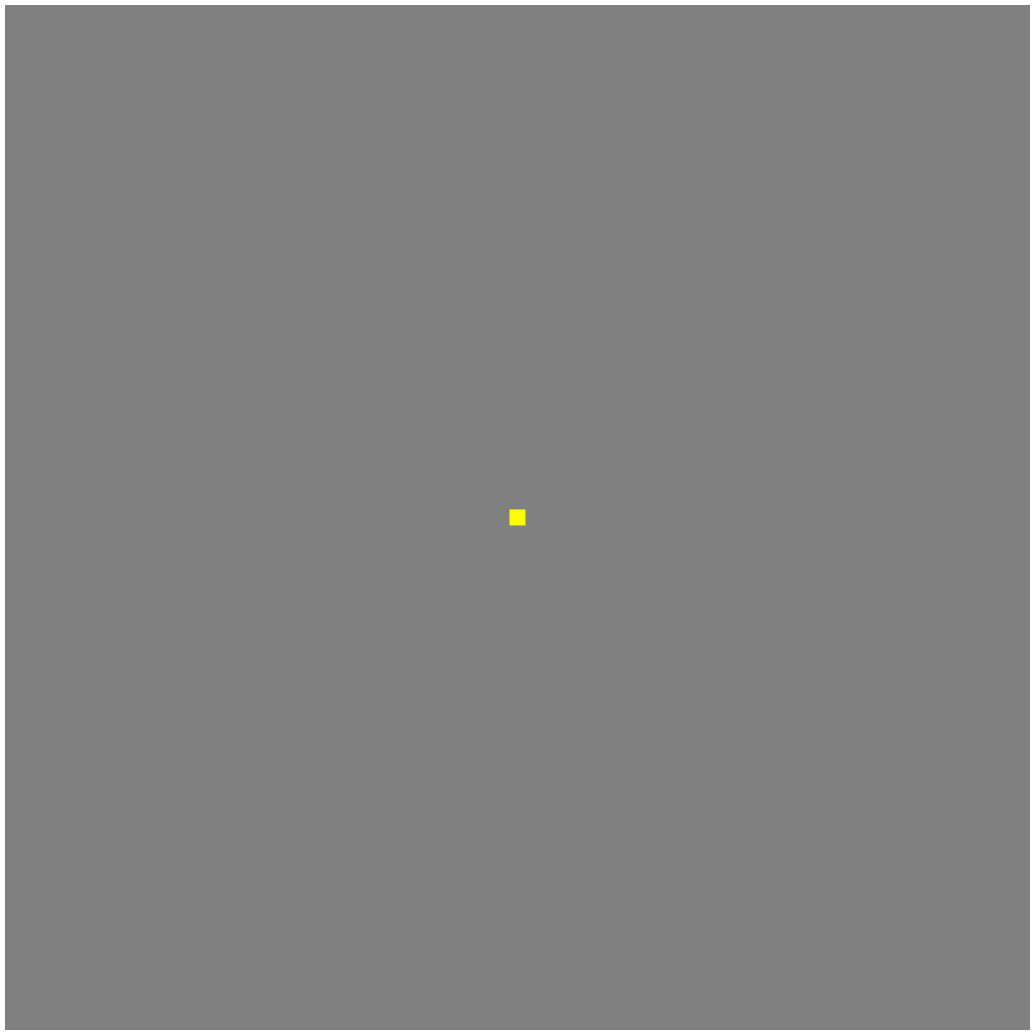
From Biederman et al., (1988)

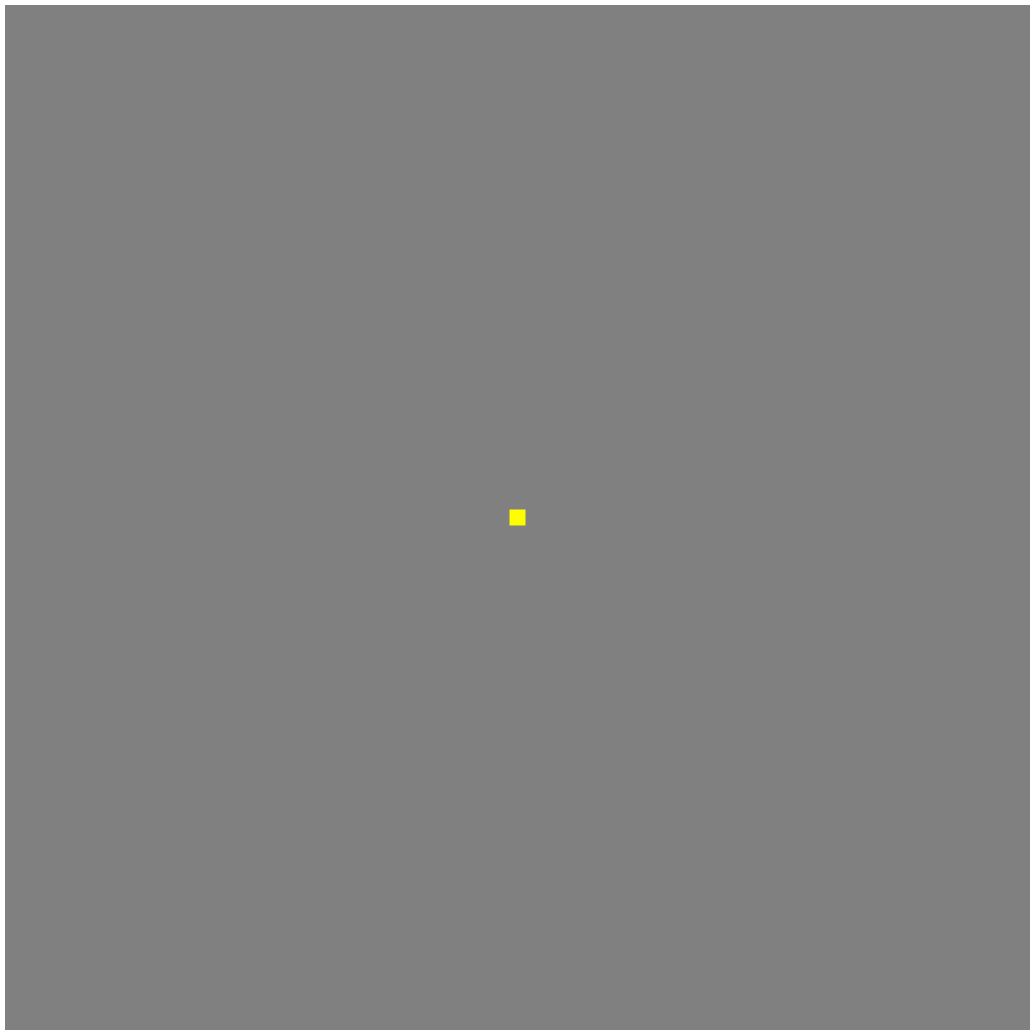
Minimal array of a pair of unrelated objects

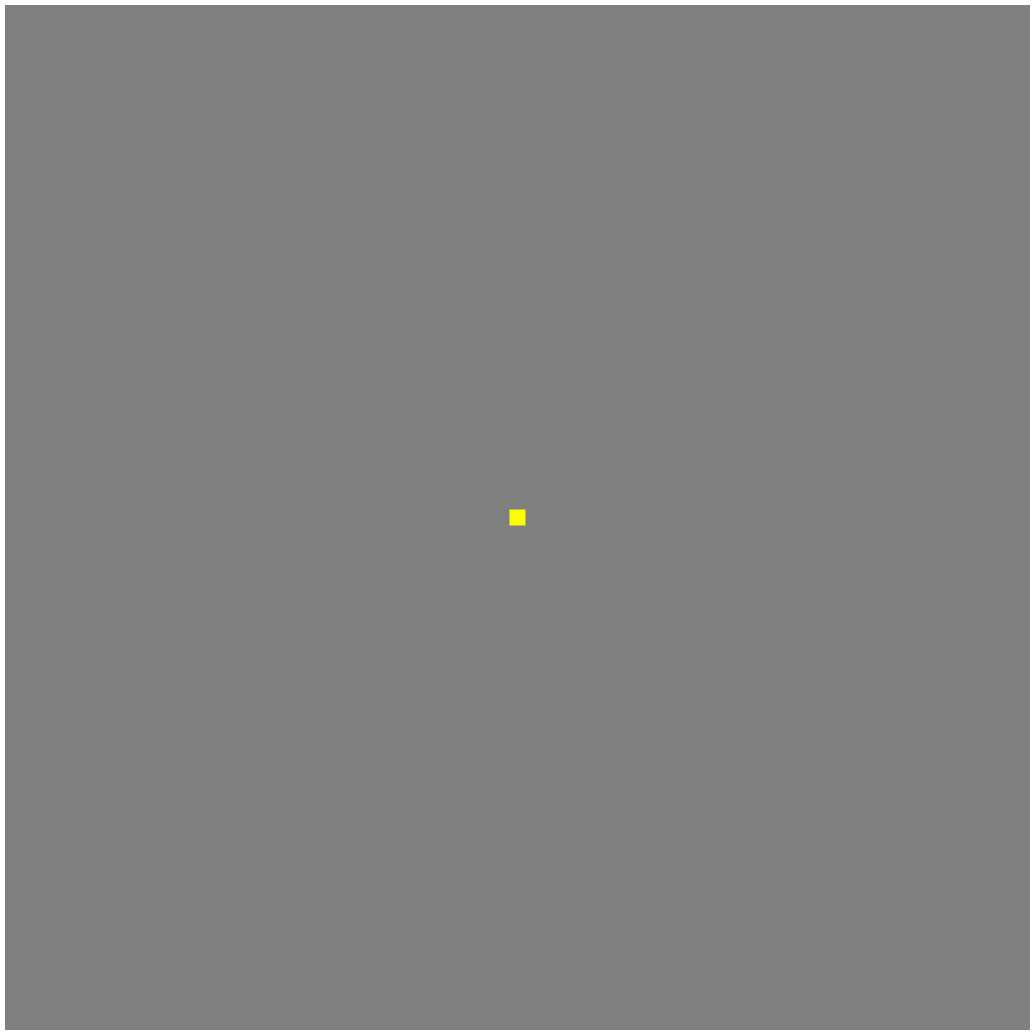
- **Is there sensitivity to a pair of unrelated objects when the only relationship is spatial proximity even when the relations are unfamiliar and not readily described by a preposition or gerund?**
- **Is this sensitivity reflected in LOC?**

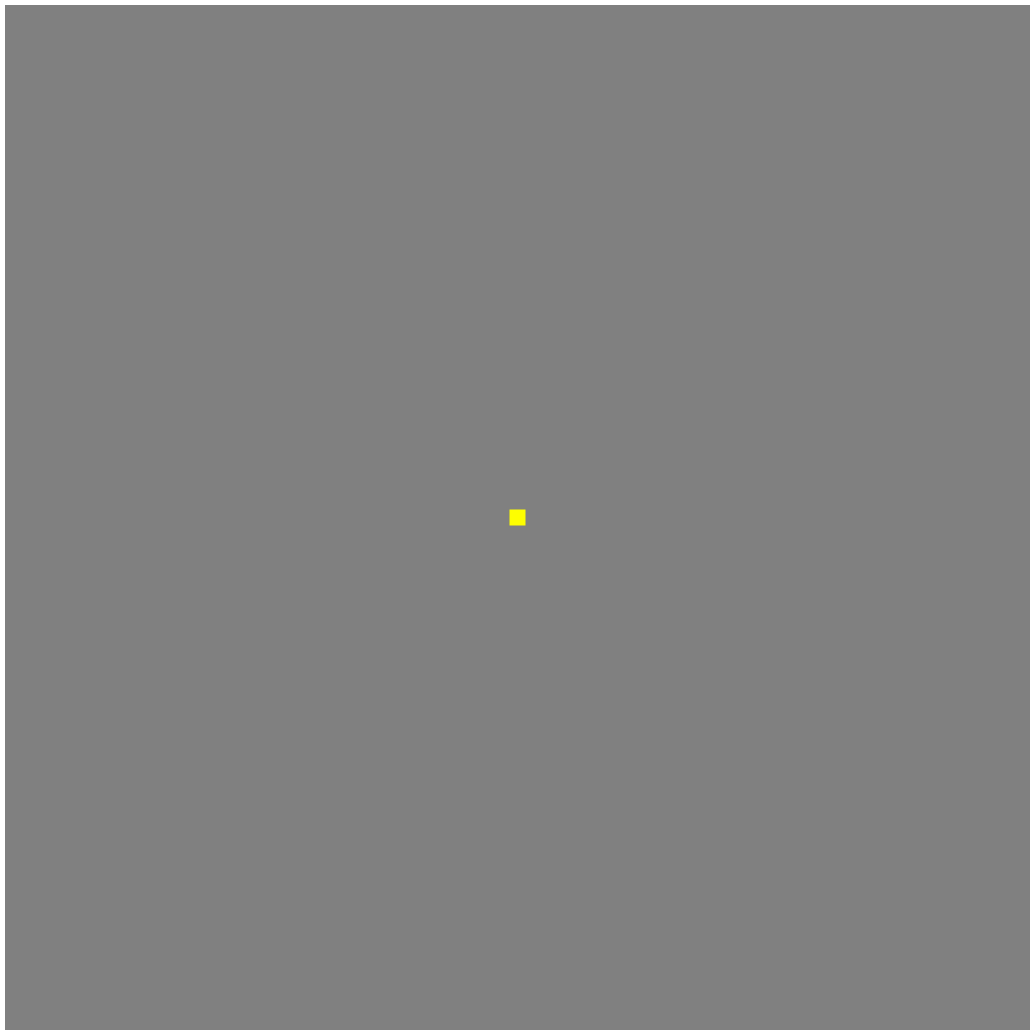

Five Trial Types
Identical
Translated
Relation
Translated + Relation
New Object
S1
S2
or…
or…
or…
Example "Relation" trial
200ms
300ms
200ms
Eight possible S1 arrangements

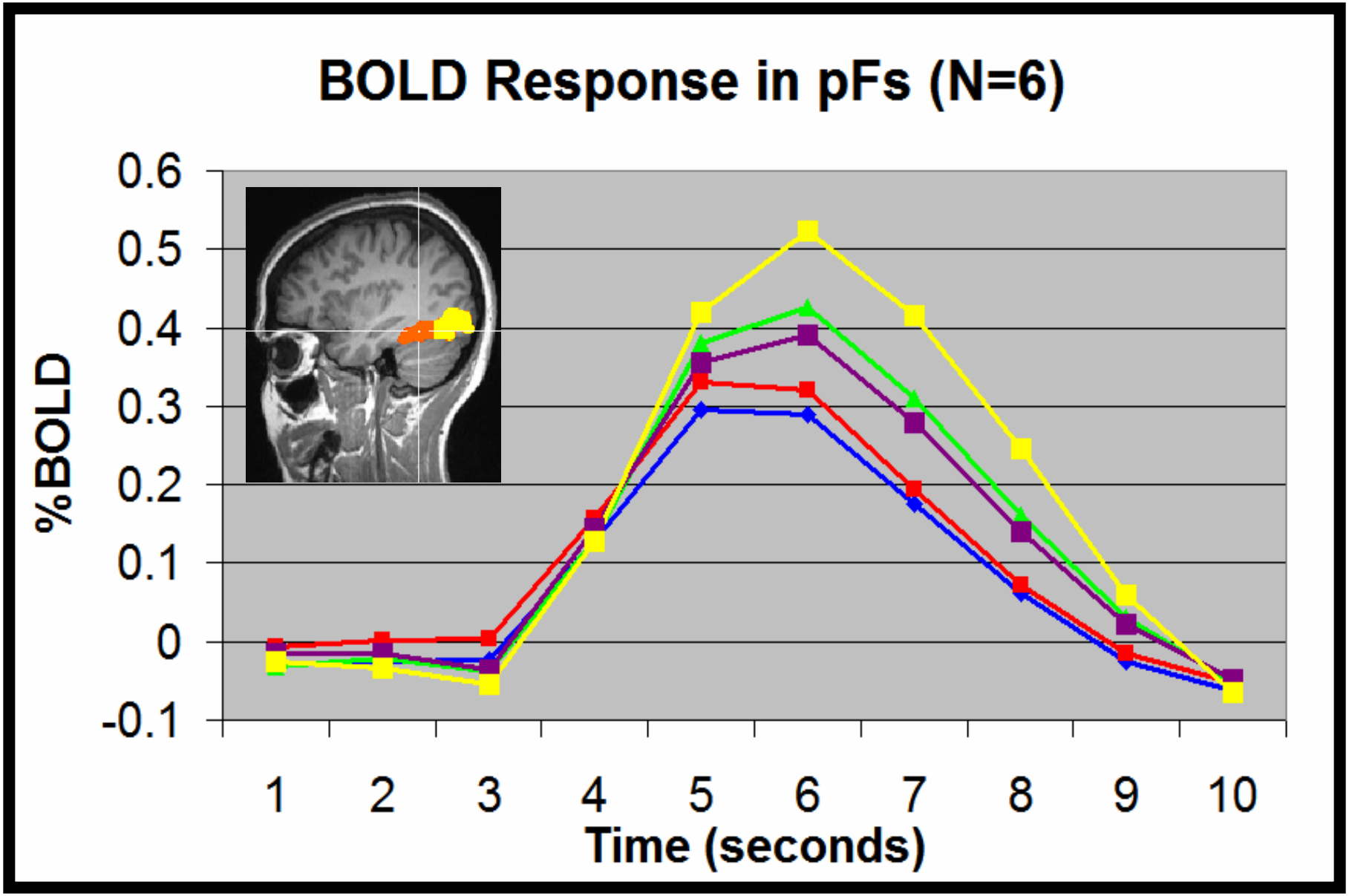




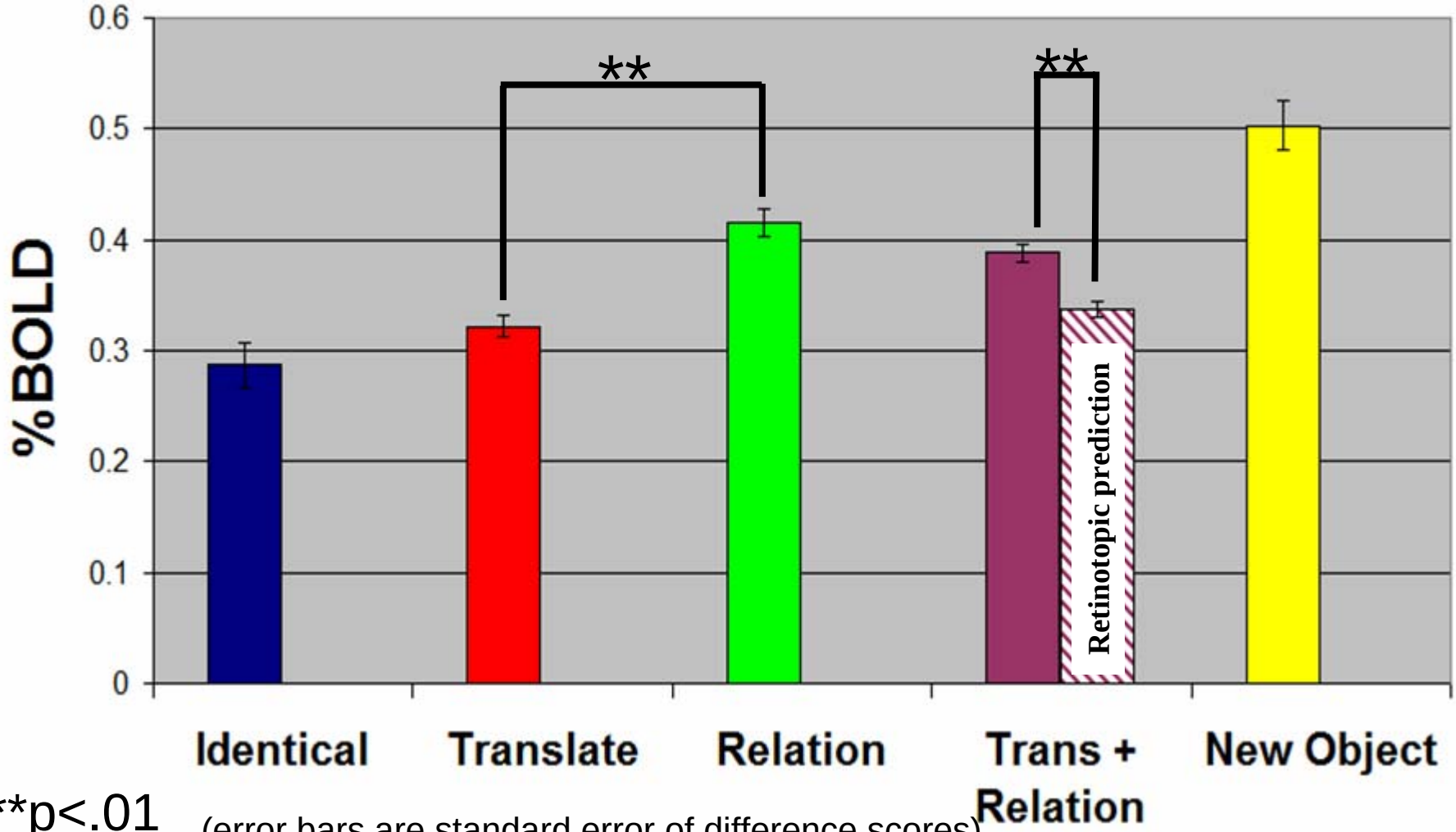







BOLD Response in pFs (N=6)
%BOLD
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
1
2
3
4
5
6
7
8
9
10
Time (seconds)

Peak BOLD Response in pFs (N=6)



**p<.01 (error bars are standard error of difference scores)

Role of familiarity of relationship between a pair of objects and their recognition in minimal scenes?

Interacting pairs (Fig. 4) produced much better cued recall than non-interacting images (Fig. 3).

Epstein, Rock, & Zuckerman (1960).

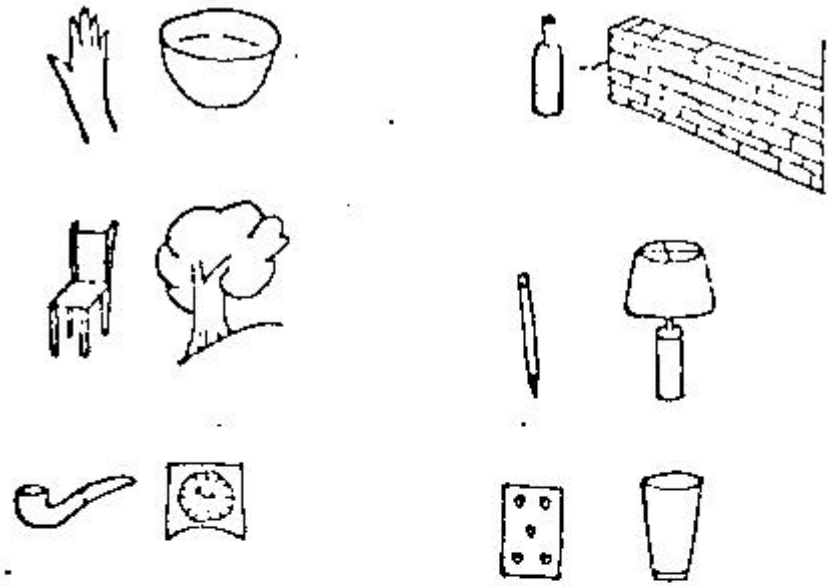


FIG. 3. Pairs learned by Group 2, Experiment VIIb.

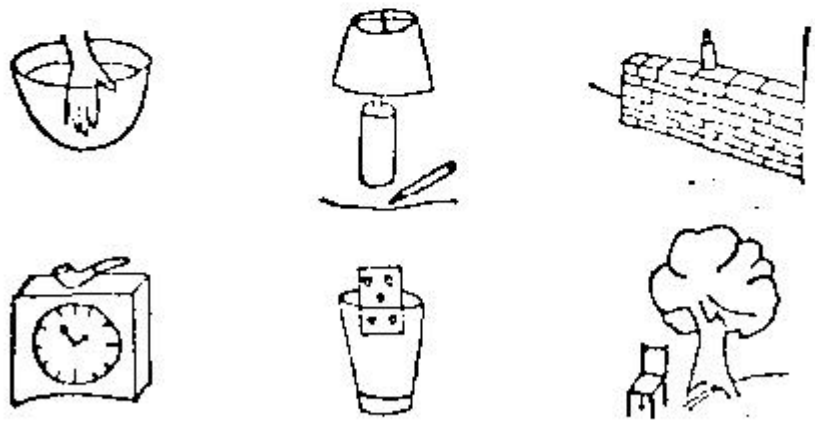
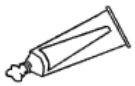
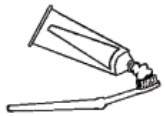
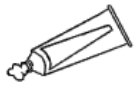


FIG. 4. Pairs learned by Group 1, Experiment VIIb.

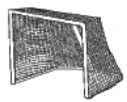
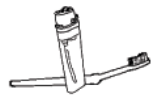
TARGET PRESENT

SINGLETON		PAIR	
Familiar	Novel	Familiar	Novel
toothpaste 	hammer & nail 	toothbrush & toothpaste 	key 
 kettle & cup	 toothpaste	 goalie	 toothpaste & pa

Target: Toothpaste

Mate: Toothbrush

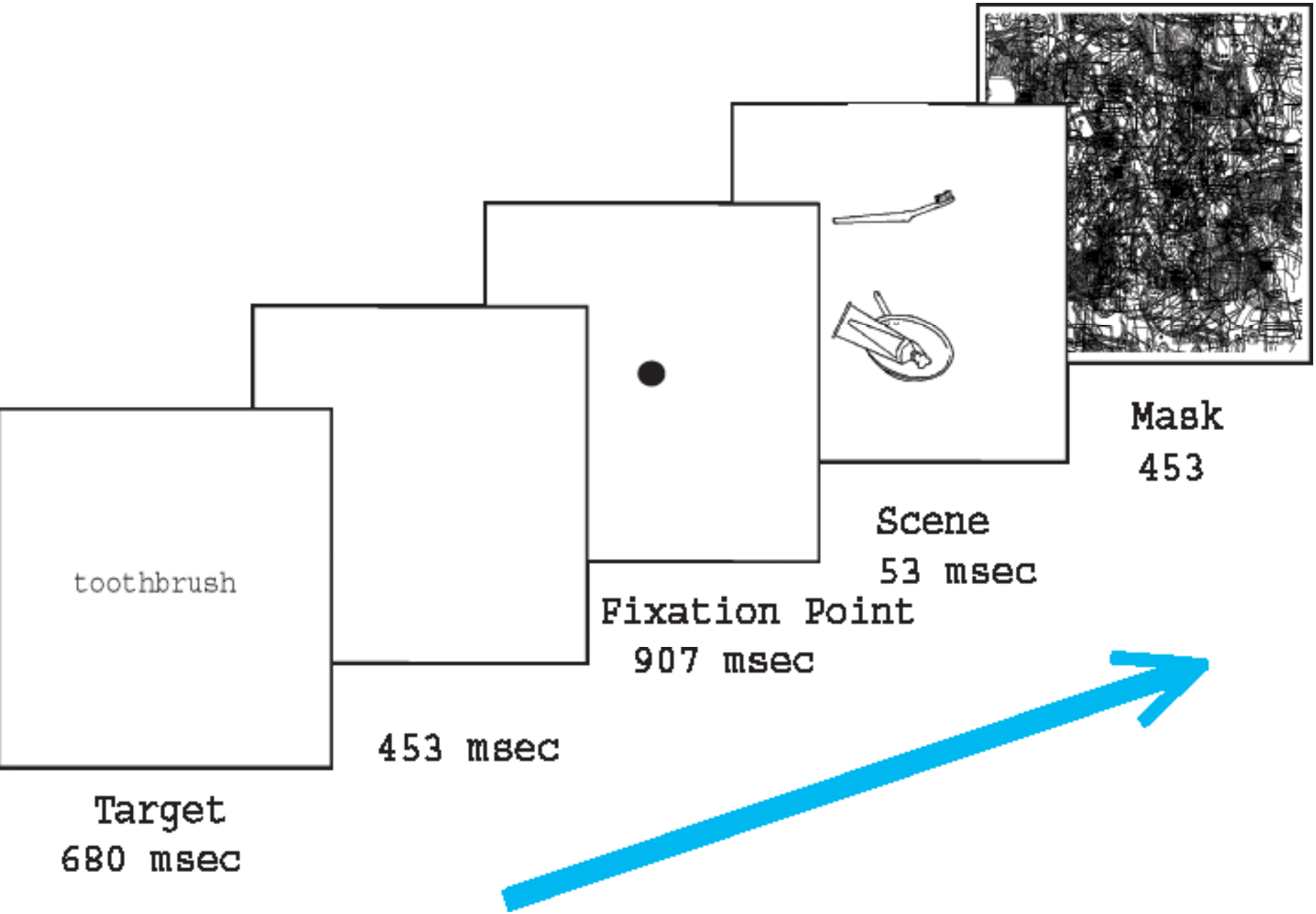
TARGET ABSENT

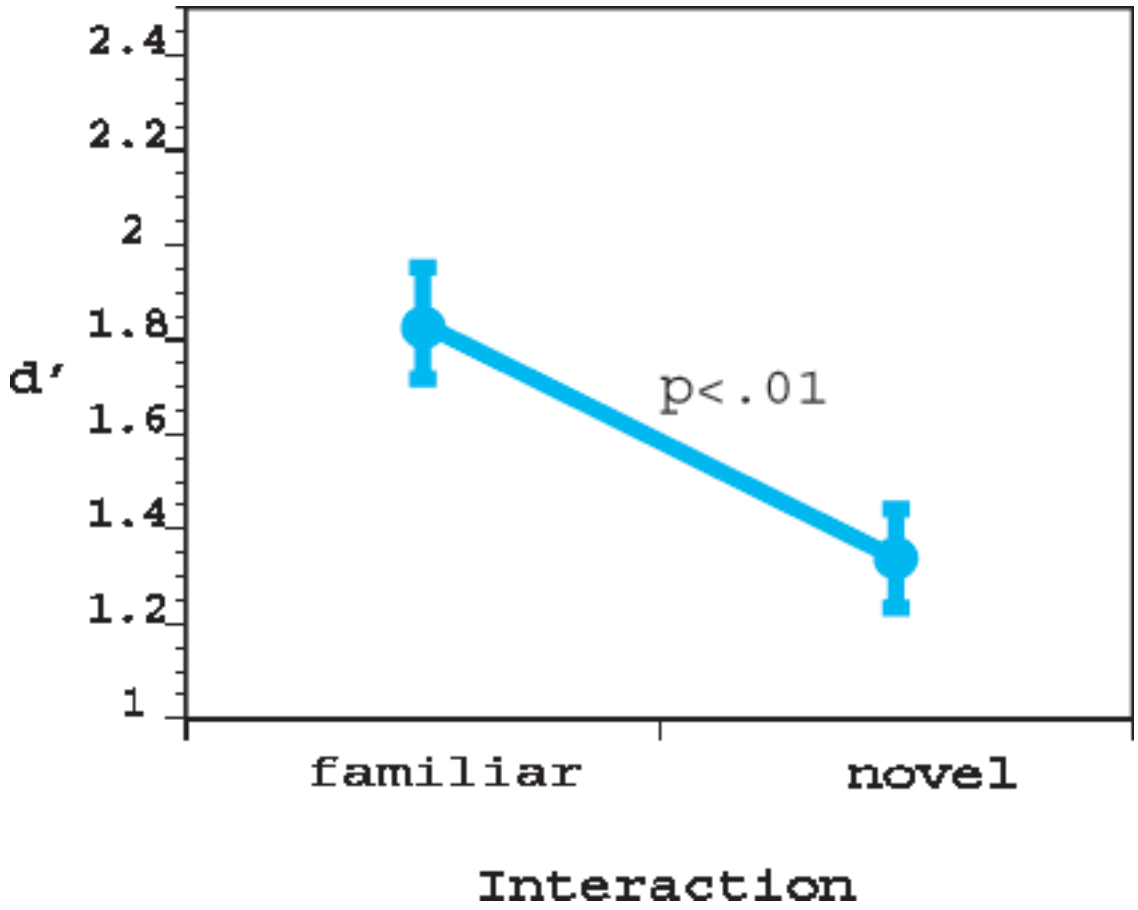
NO MATE		MATE	
Familiar	Novel	Familiar	Novel
car & gas pump 	goal 	toothbrush 	toothbrush & lighter 
 flower	 hand & flower	 man & electric razor	 lock


QuickTime™ and a
TIFF (Uncompressed) decompressor
are needed to see this picture.


QuickTime™ and a
TIFF (Uncompressed) decompressor
are needed to see this picture.


QuickTime™ and a
TIFF (Uncompressed) decompressor
are needed to see this picture.


toothbrush
Target
680 msec
453 msec
Fixation Point
907 msec
Scene
53 msec
Mask
453


2.4
2.2
2
1.8
1.6
1.4
1.2
1
d'
p<.01
familiar
novel
Interaction

Role of Familiarity of Relationships in Complex Scenes

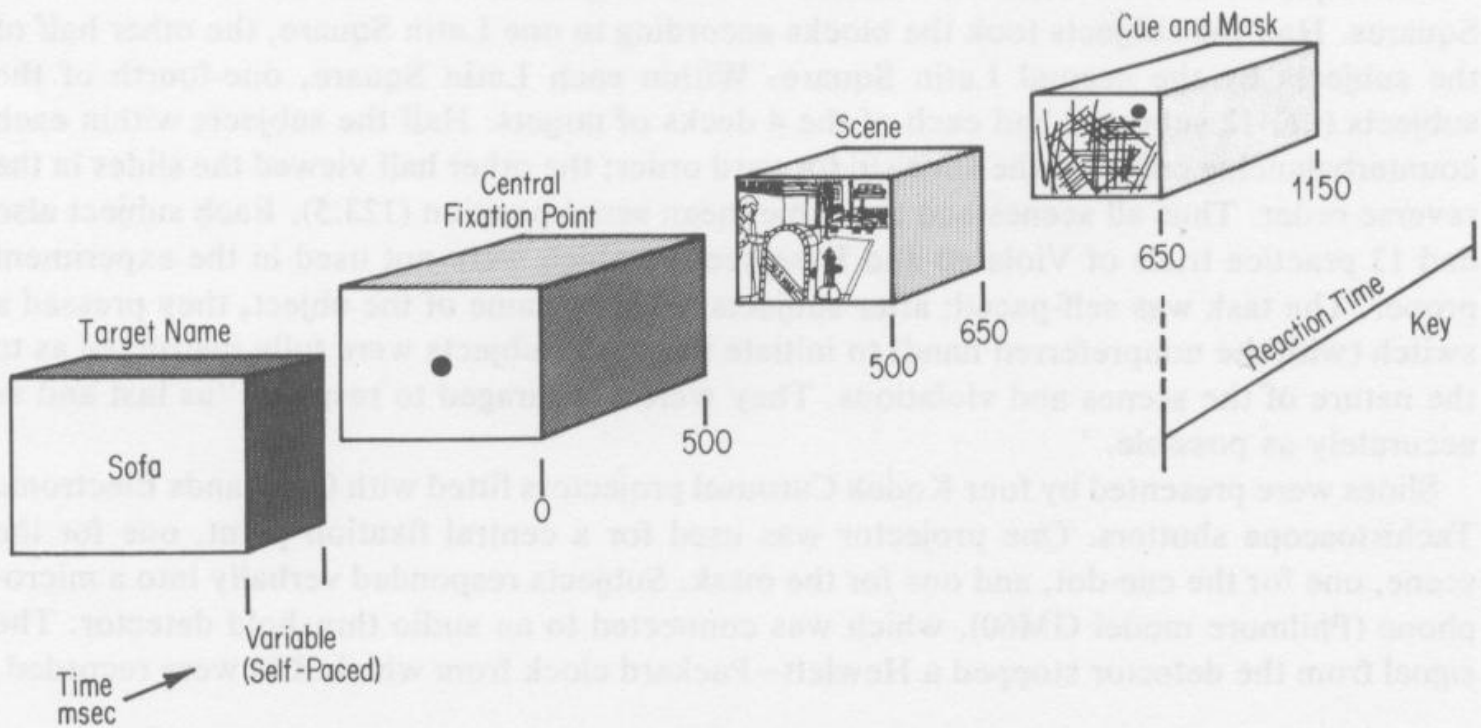
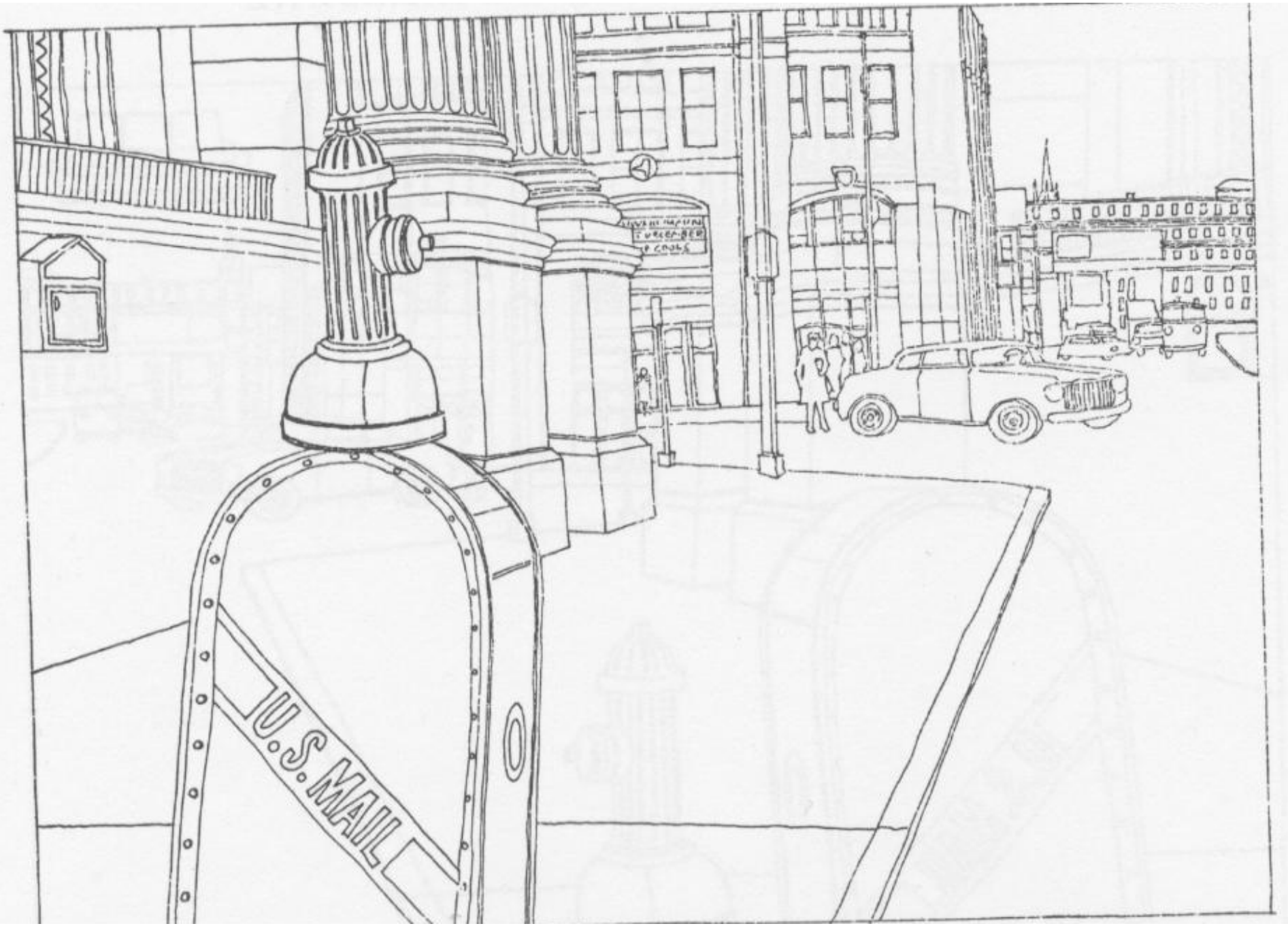


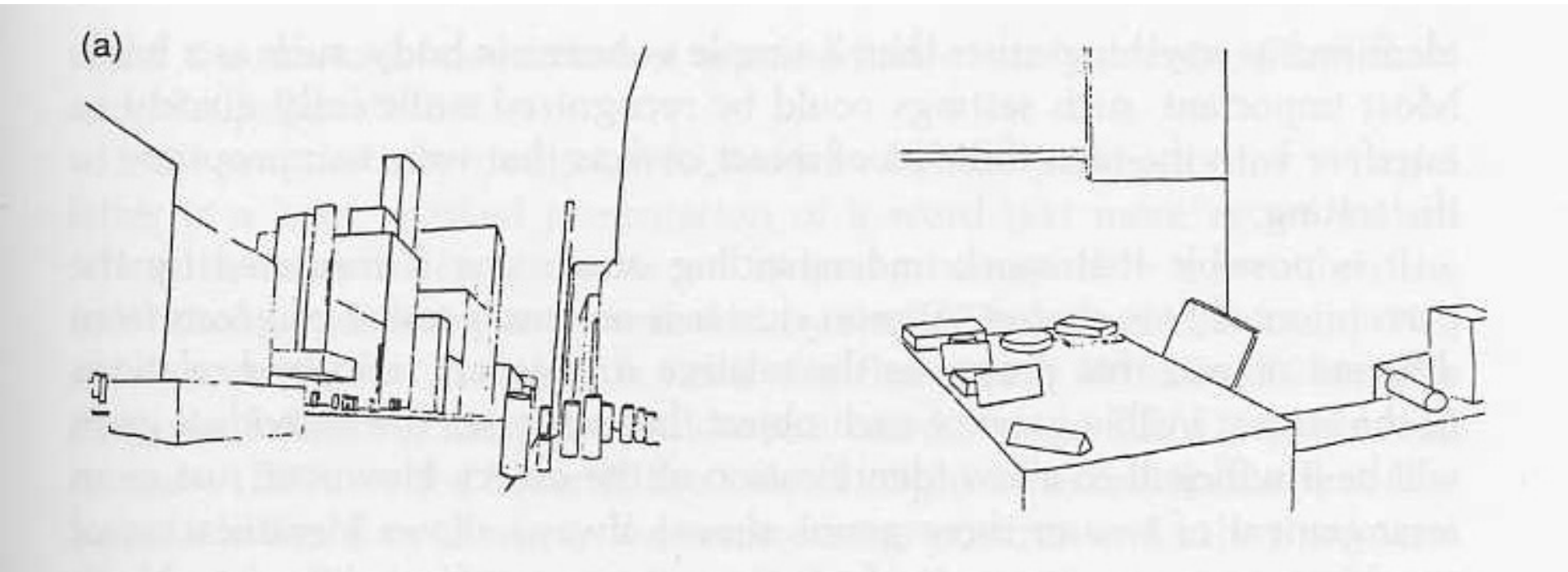
FIG. 4. Sequence of events in the Object Detection task.

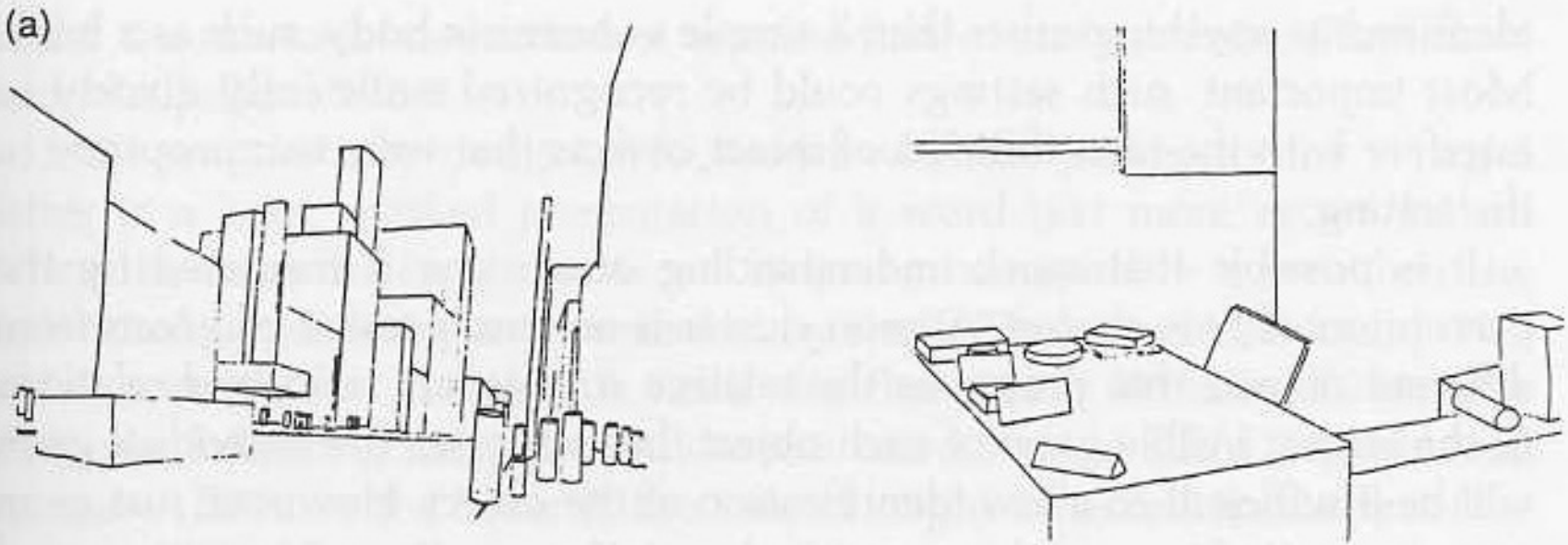
Biederman, Mezzonotte, Klatsky, & Teitelbaum, 1982


U.S. MAIL

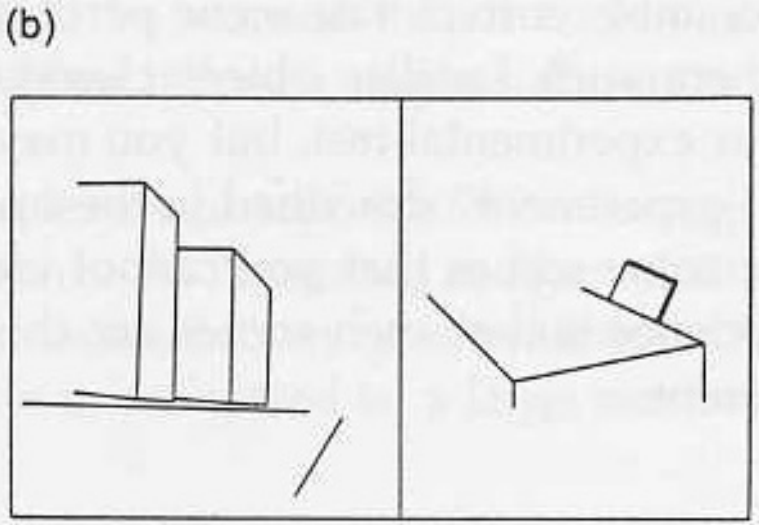

QuickTime™ and a
TIFF (Uncompressed) decompressor
are needed to see this picture.

Well formed objects are not required for rapid setting classification!





Setting classification may be mediated by learned *clusters* (or "*scenelets*") of simple shapes. Humans may have stored thousands (if not tens of thousands)



Scenelets (formerly geon clusters) but same idea as parts-based object representation.

Setting Verification vs. Scene Understanding

Setting Verification (Similar vs. Dissimilar Descriptors) in Jumbled vs. Intact Scenes

Example of Similar Descriptors: City Street vs. Parking Lot
Example of Dissimilar Descriptors: City Street vs. Restaurant Kitchen

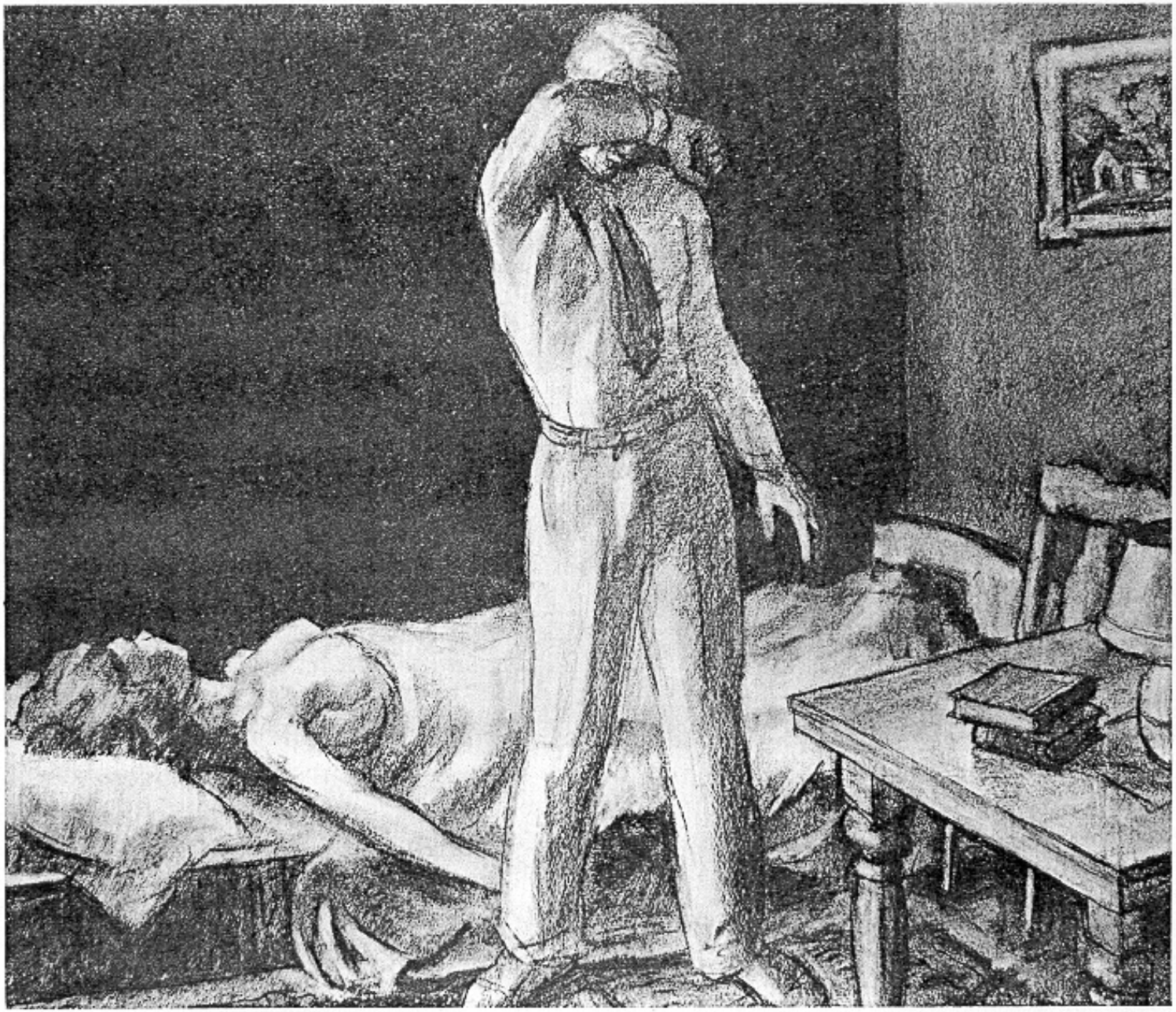
QuickTime™ and a
TIFF (Uncompressed) decompressor
are needed to see this picture.

Biederman, Rabinowitz, Glass, & Stacy (1974)

Railroad Station


CHEMIN DE FER
DE L'OUEST
RESTAURANT

Bedroom



Happy Tourists

Difficulty in Measuring Scene Interpretation

One method:

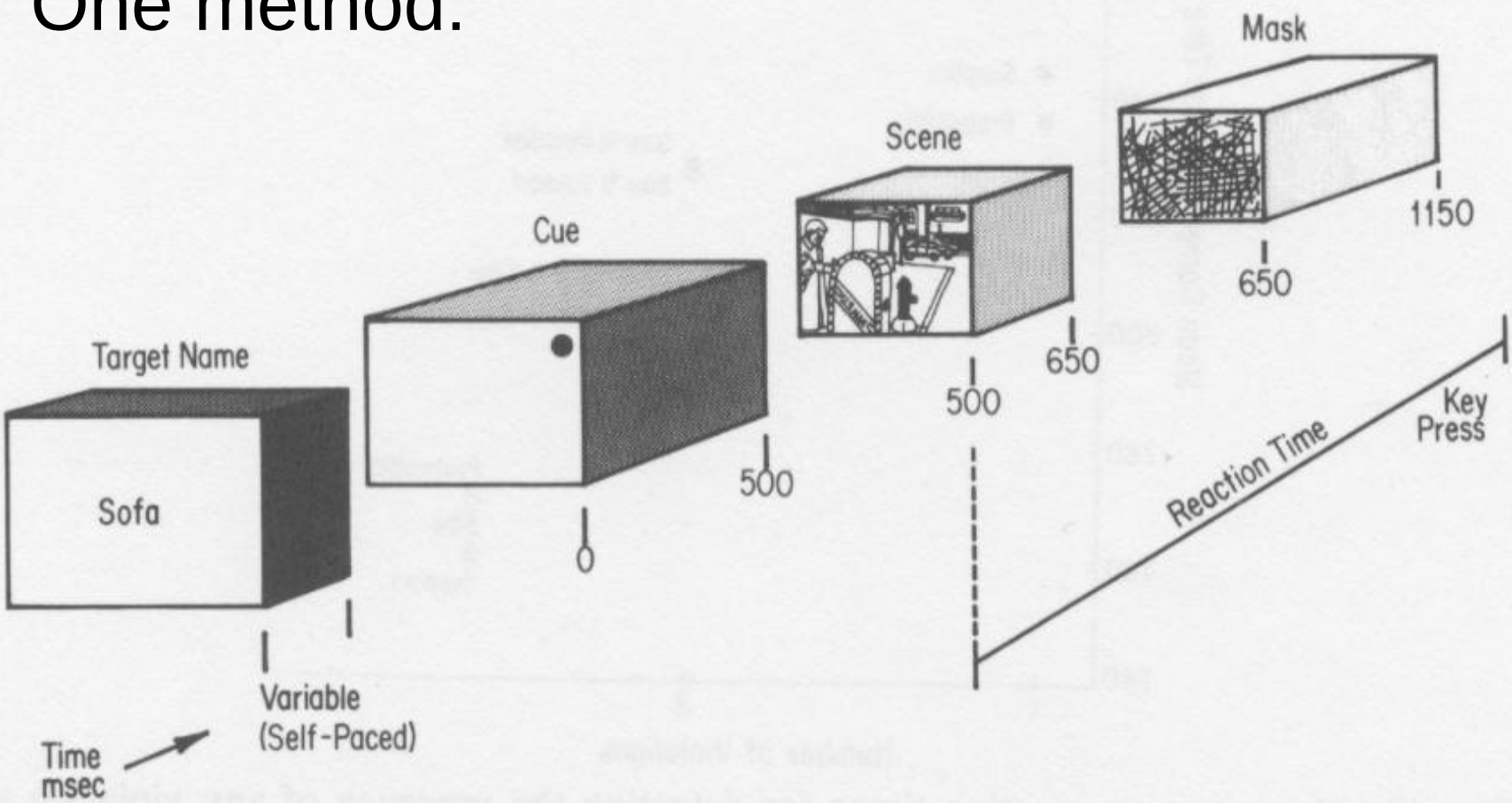


FIG. 12. Sequence of events in the Violation Detection task.

The Grand Challenge of Visual Cognition

- **How can we achieve scene understanding (not just setting classification)?**
- **In trying to understand how we can so quickly understand novel scenes, we face many of the same daunting problems faced by linguists and psycholinguists in trying to account for the representations activated by discourse.**