The Perceptual Representation of 3D Shape

James T. Todd

Ohio State University

Collaborators

Andrea Van Doorn

Astrid Kappers

Jan Koenderink

Joe Lappin

Baoxia Liu

Farley Norman

Stijn Oomes

Victor Perotti

Flip Phillips

The Physical Environment Φ

$\Lambda = f(\Phi)$

The Structure of Light Λ

Physiological Processes

Mathematical Algorithms

Prior Knowledge

A Miracle Occurs

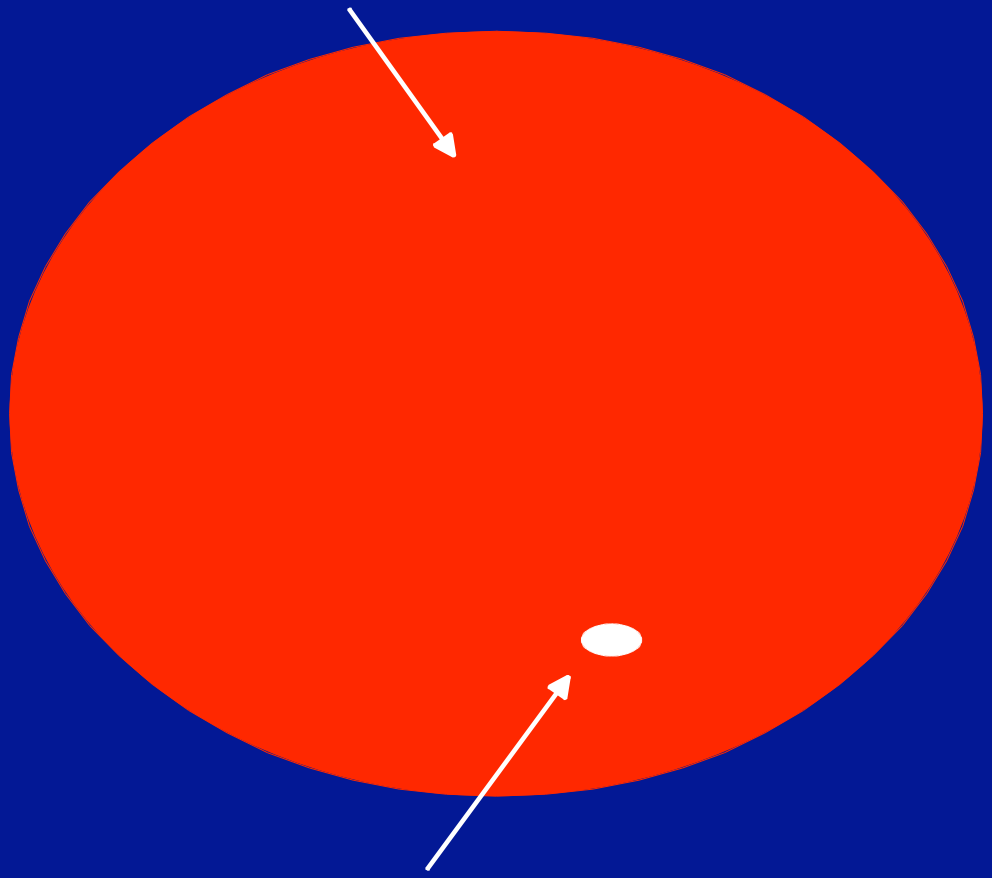
Perception of the Environment Ψ

The transformation $\Lambda = f(\Phi)$ is a many-to-one mapping

$$(X, Y, Z) \longmapsto (X', Y')$$

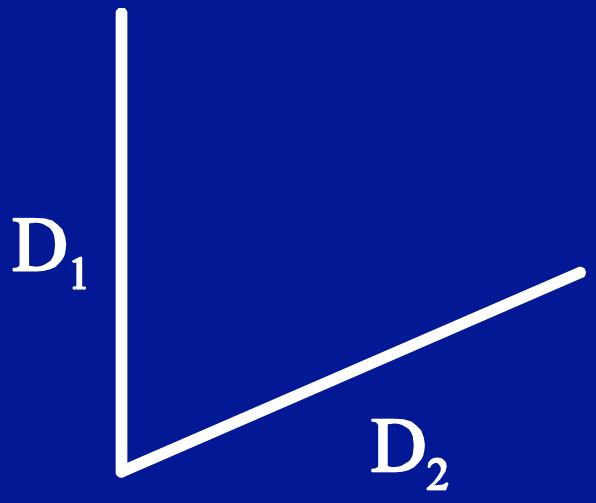
Visual information is almost always ambiguous with an infinity of possible 3D interpretations

The set of all possible 3D structures



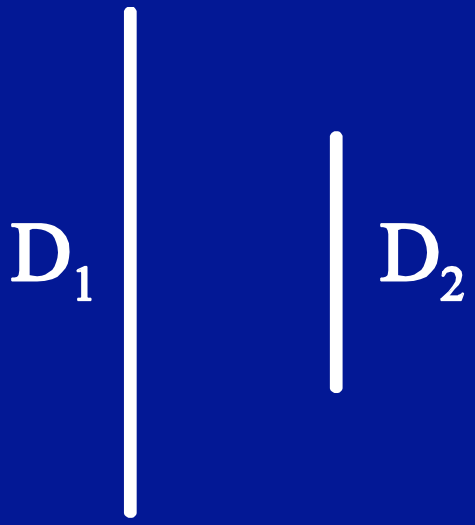
Structures that are compatible with a particular pattern of visual stimulation

Relations among distance intervals in different directions are called **Metric Structure**



Metric structure is seldom specified by patterns of visual information.

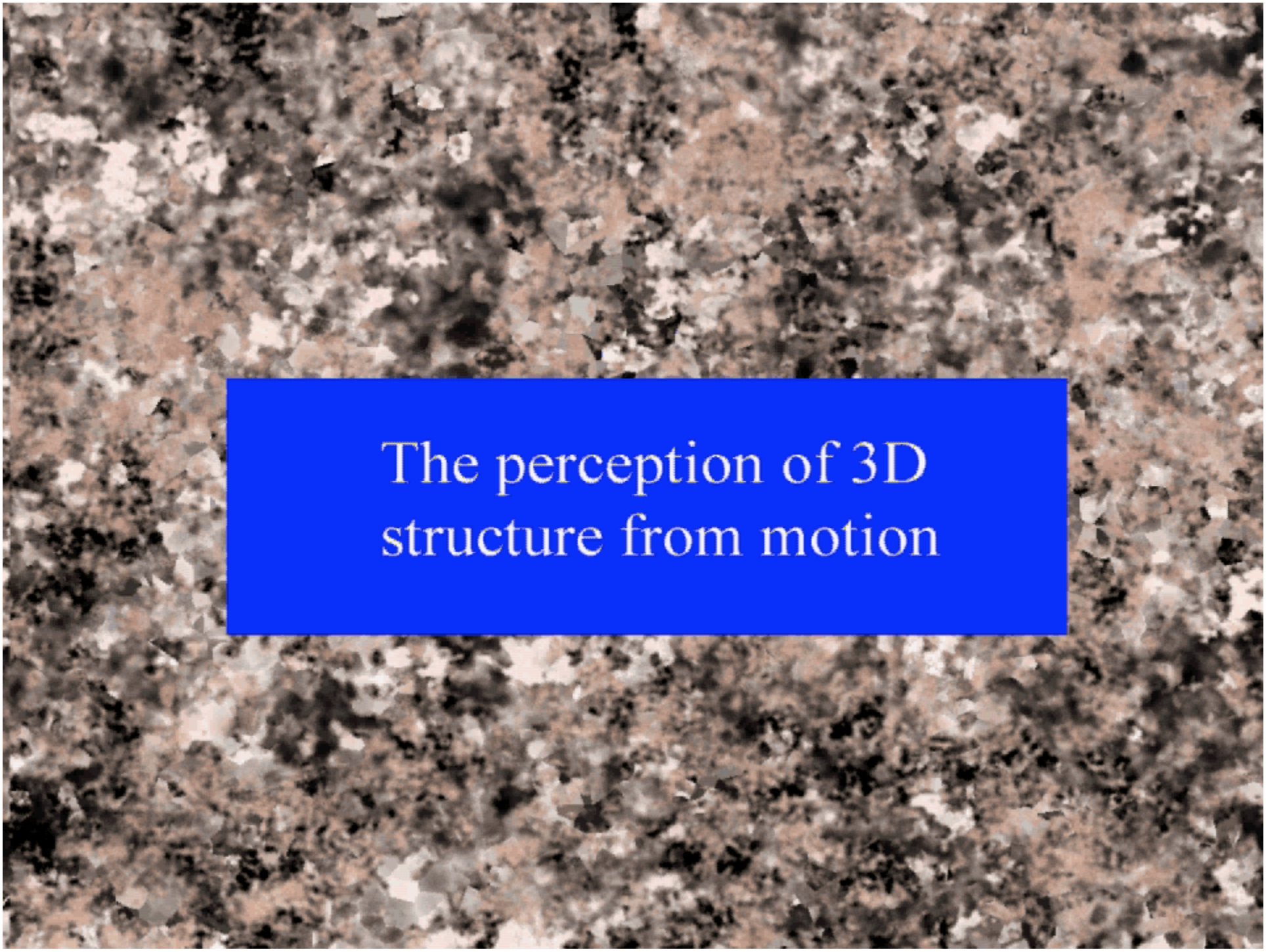
Relations among distance intervals in parallel directions are called **Affine Structure**



Affine structure is often specified by patterns of visual information.

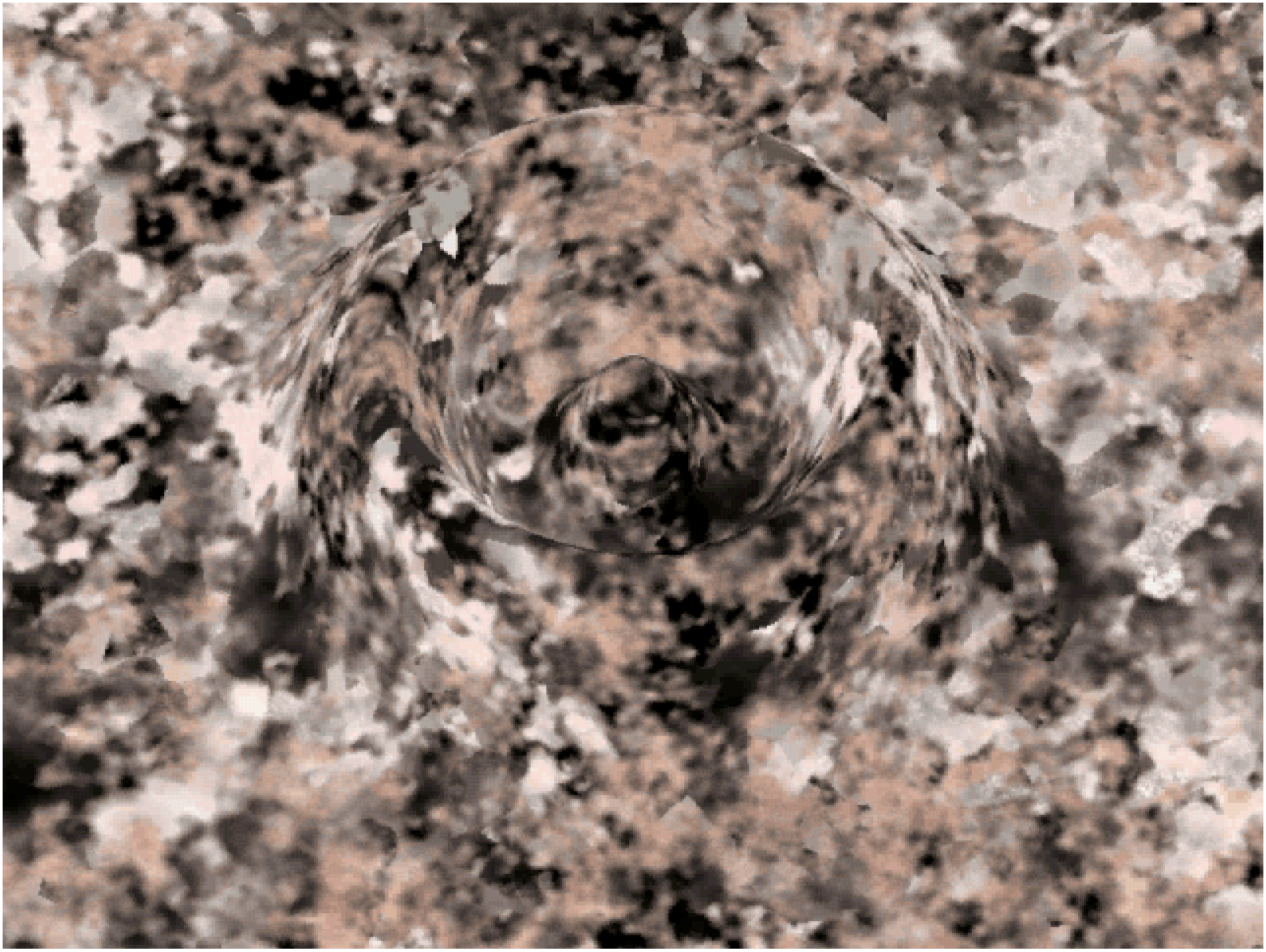
Overview of this Presentation

Source of Information	Type of Ambiguity
Motion	Depth Scaling
Texture	Depth Scaling
Shading	Depth Scaling + Shear
Binocular Disparity	Nonlinear Depth Scaling


The perception of 3D
structure from motion

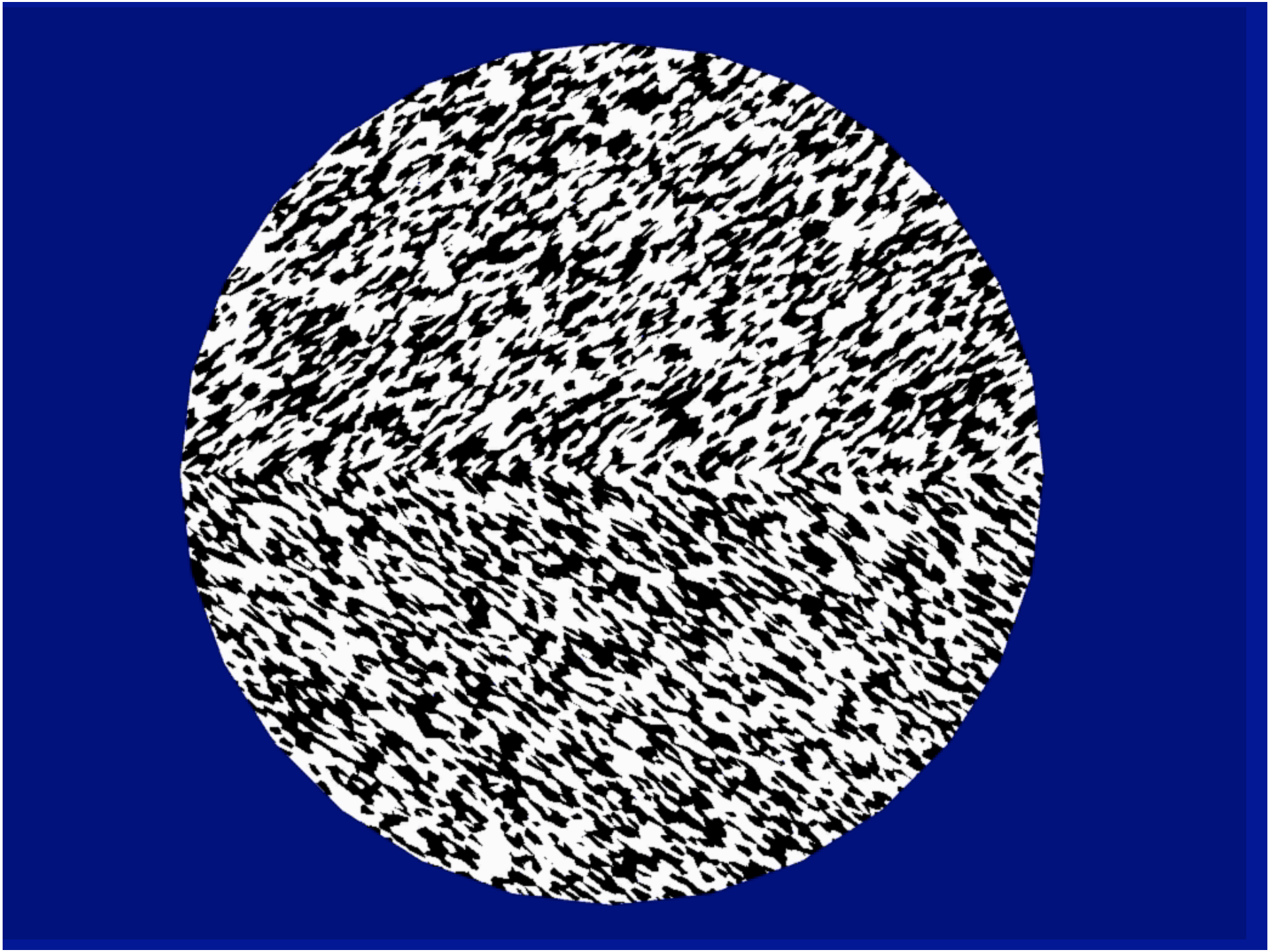
Ullman (1979)

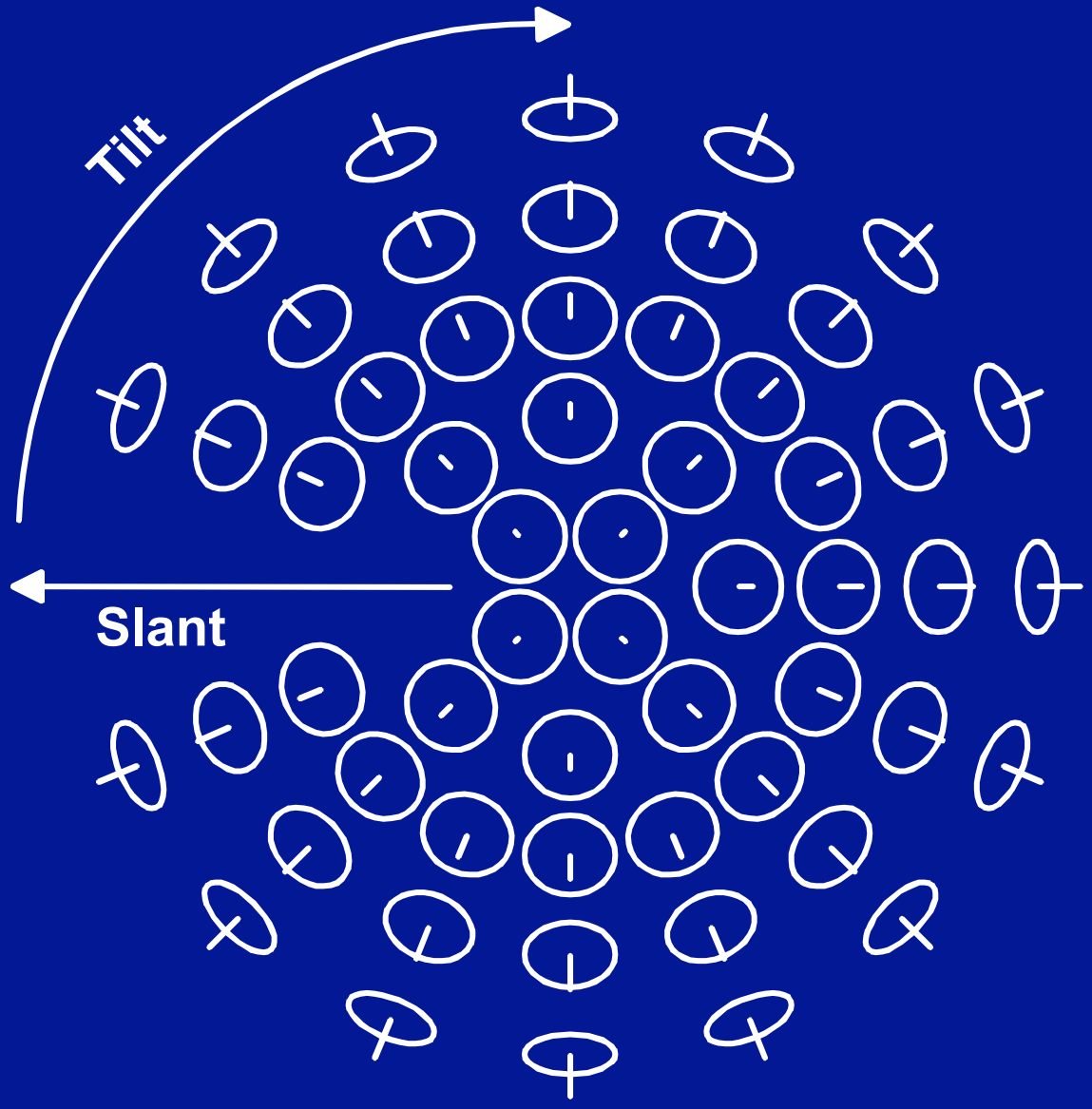
In order to determine a unique 3D structure from a sequence of images it is necessary to have at least 3 distinct views of 4 points.

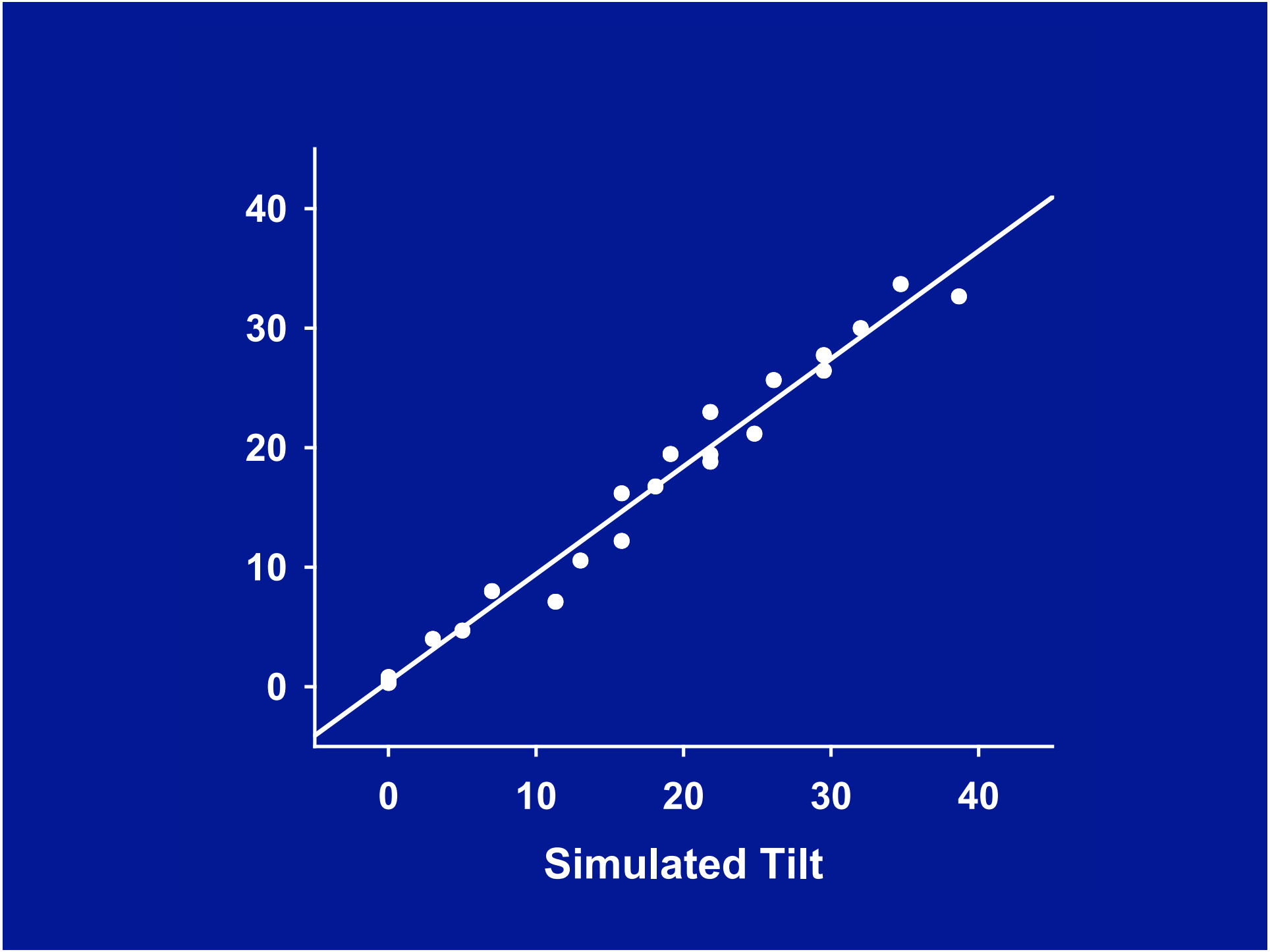


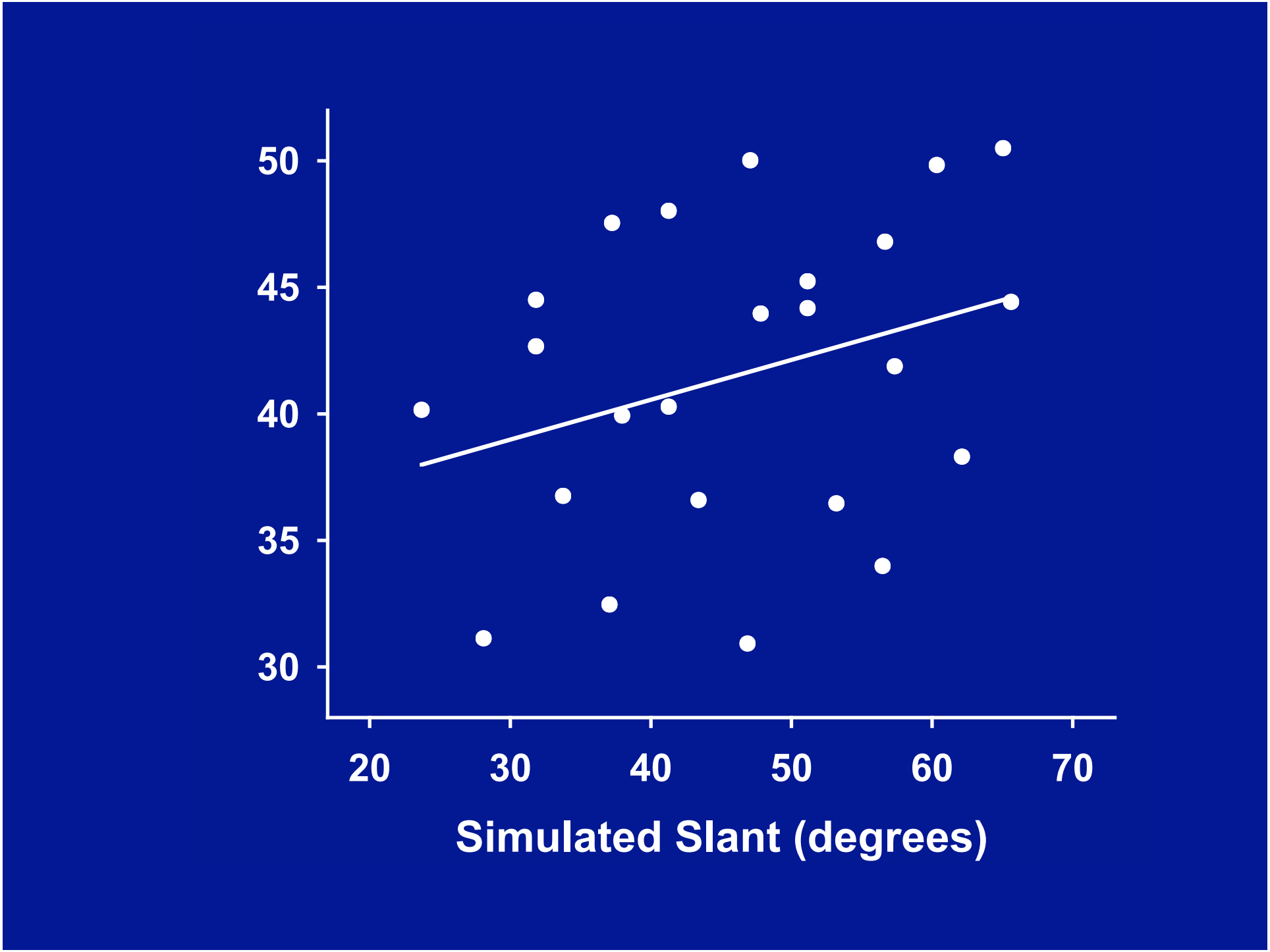
Todd & Bressan (1990)

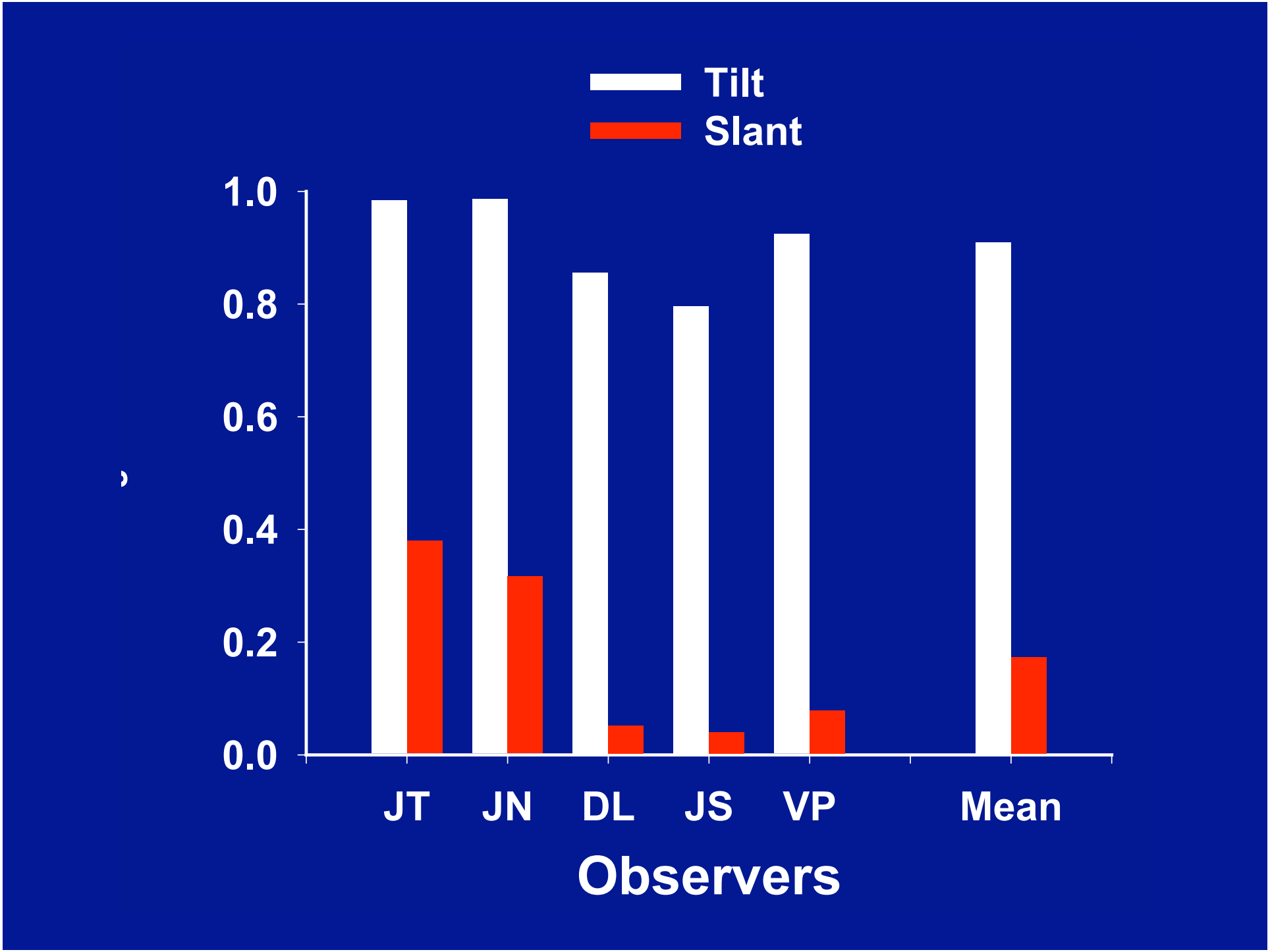
For a motion sequence composed of just two views, the 3D structure is specified up to an unknown scaling in depth.

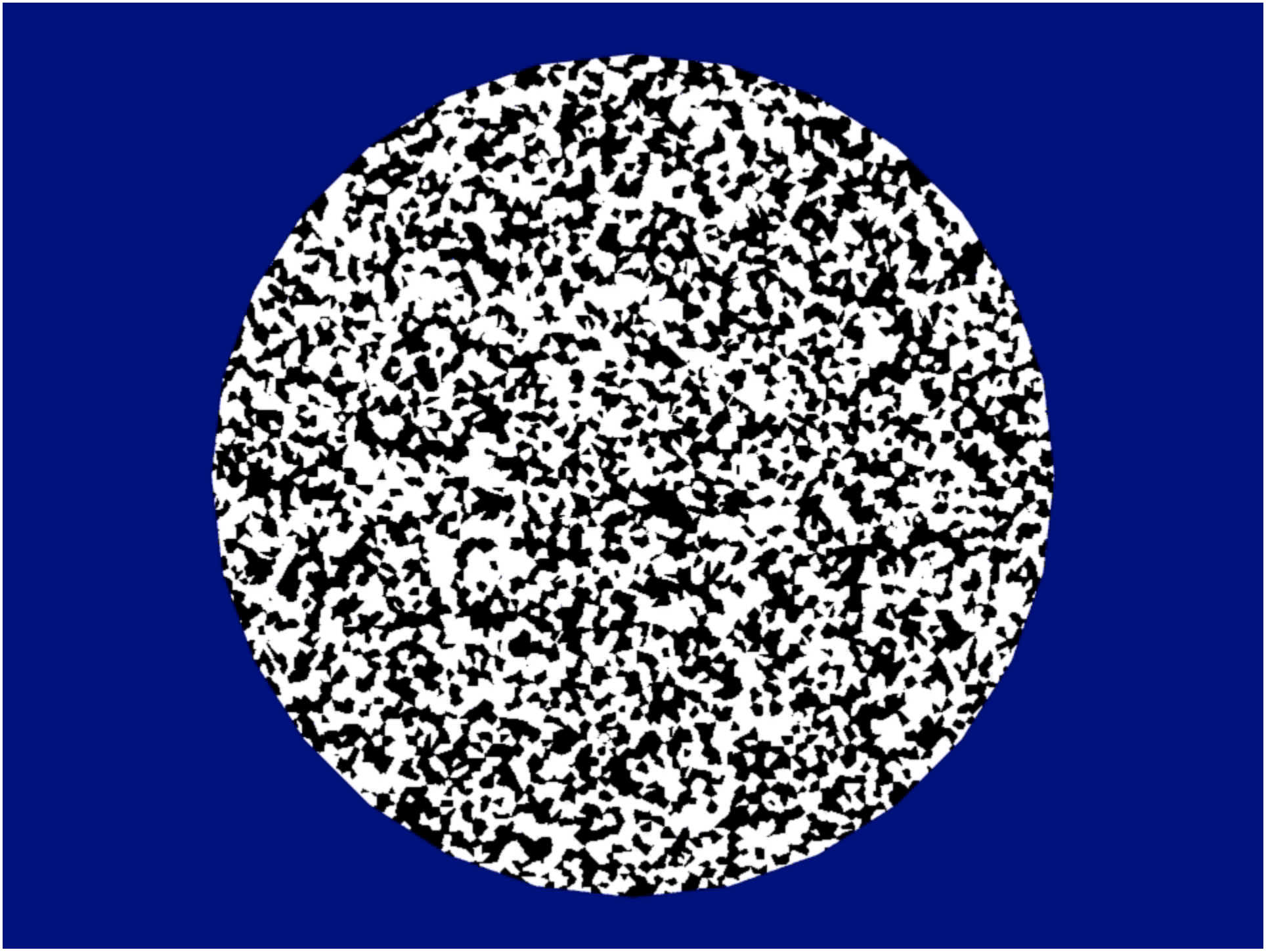


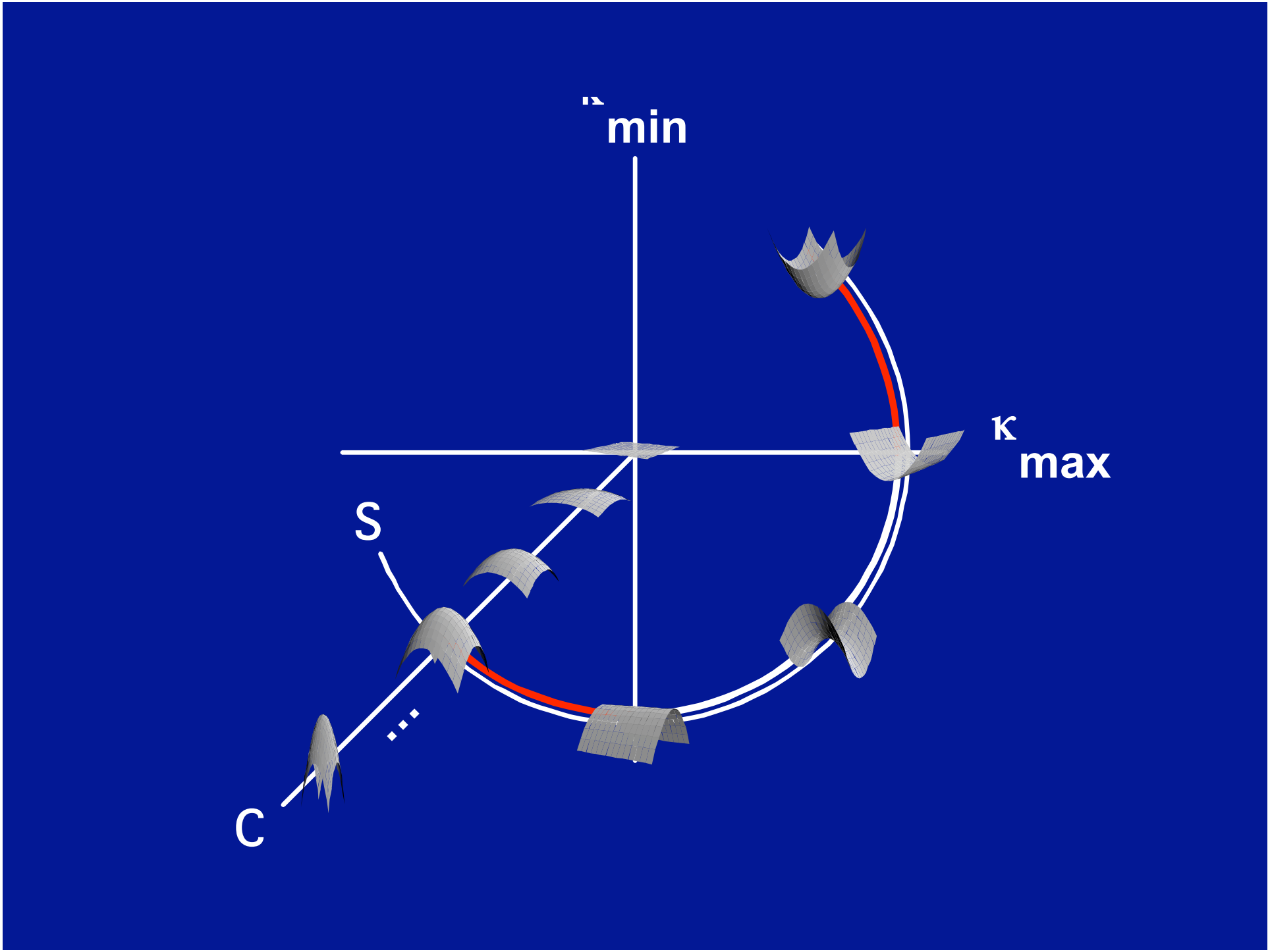

Tilt
Slant

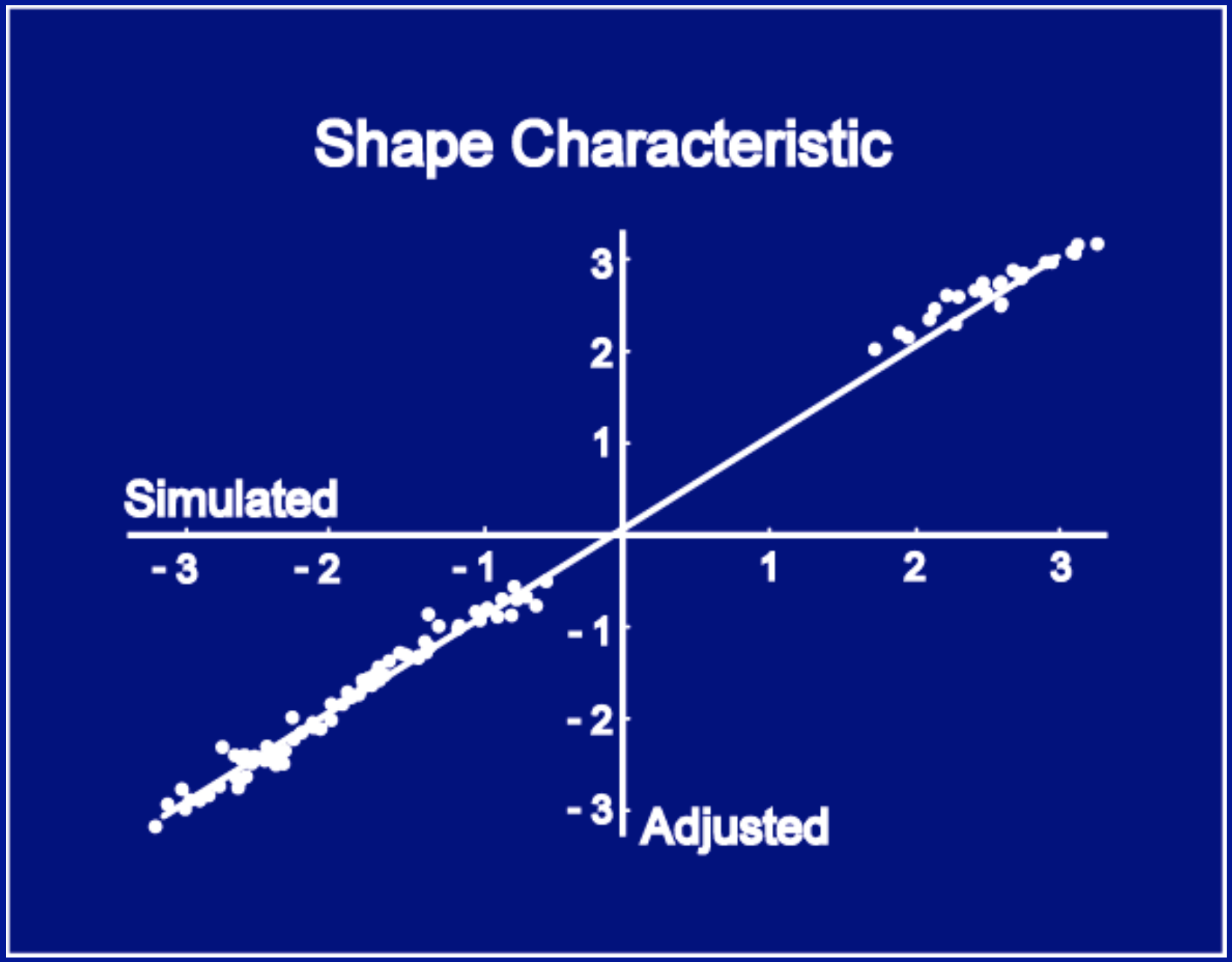

40
30
20
10
0
0
10
20
30
40
Simulated Tilt

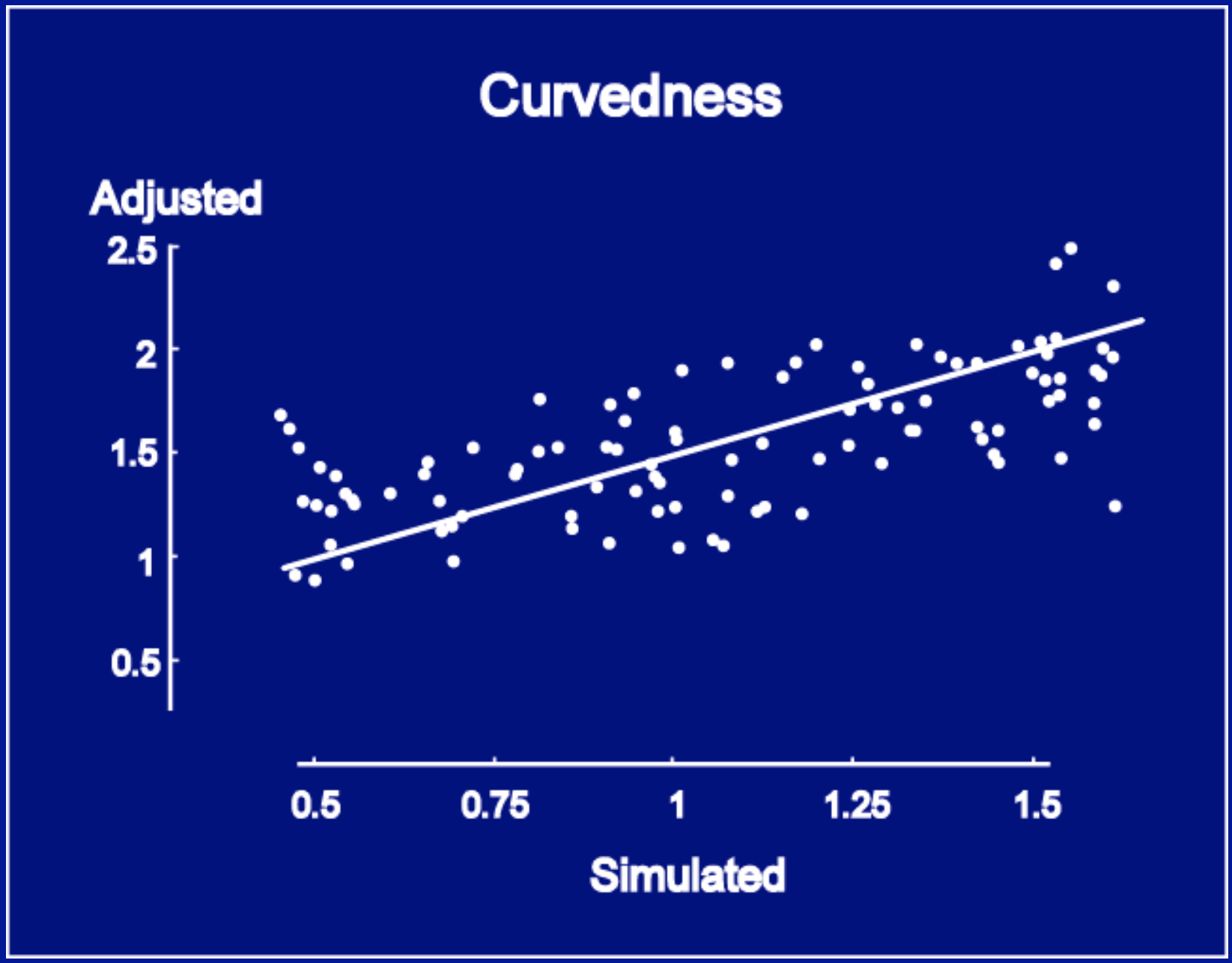

50
45
40
35
30
20
30
40
50
60
70
Simulated Slant (degrees)

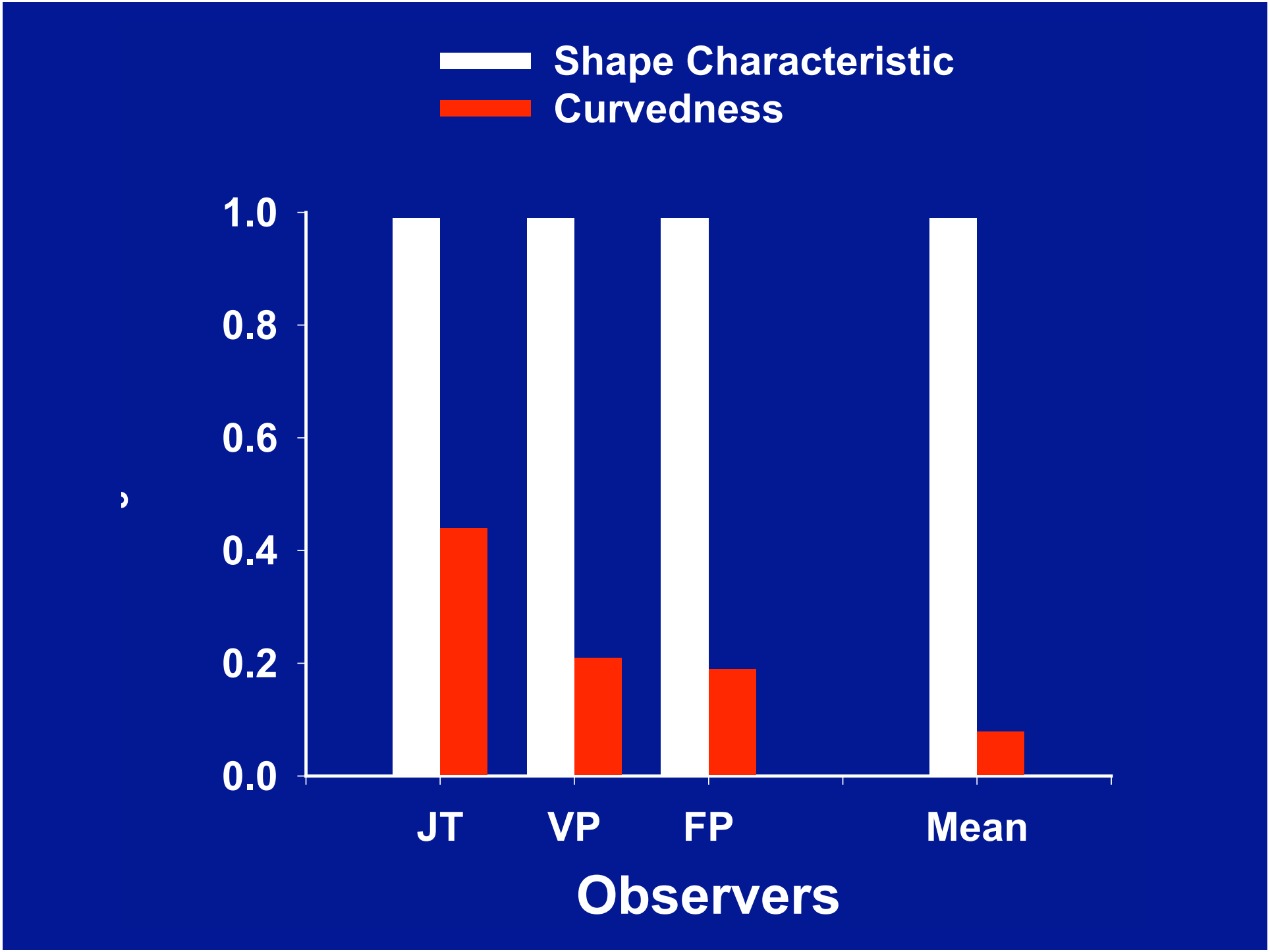

Tilt
Slant
1.0
0.8
0.6
0.4
0.2
0.0
JT
JN
DL
JS
VP
Mean
Observers



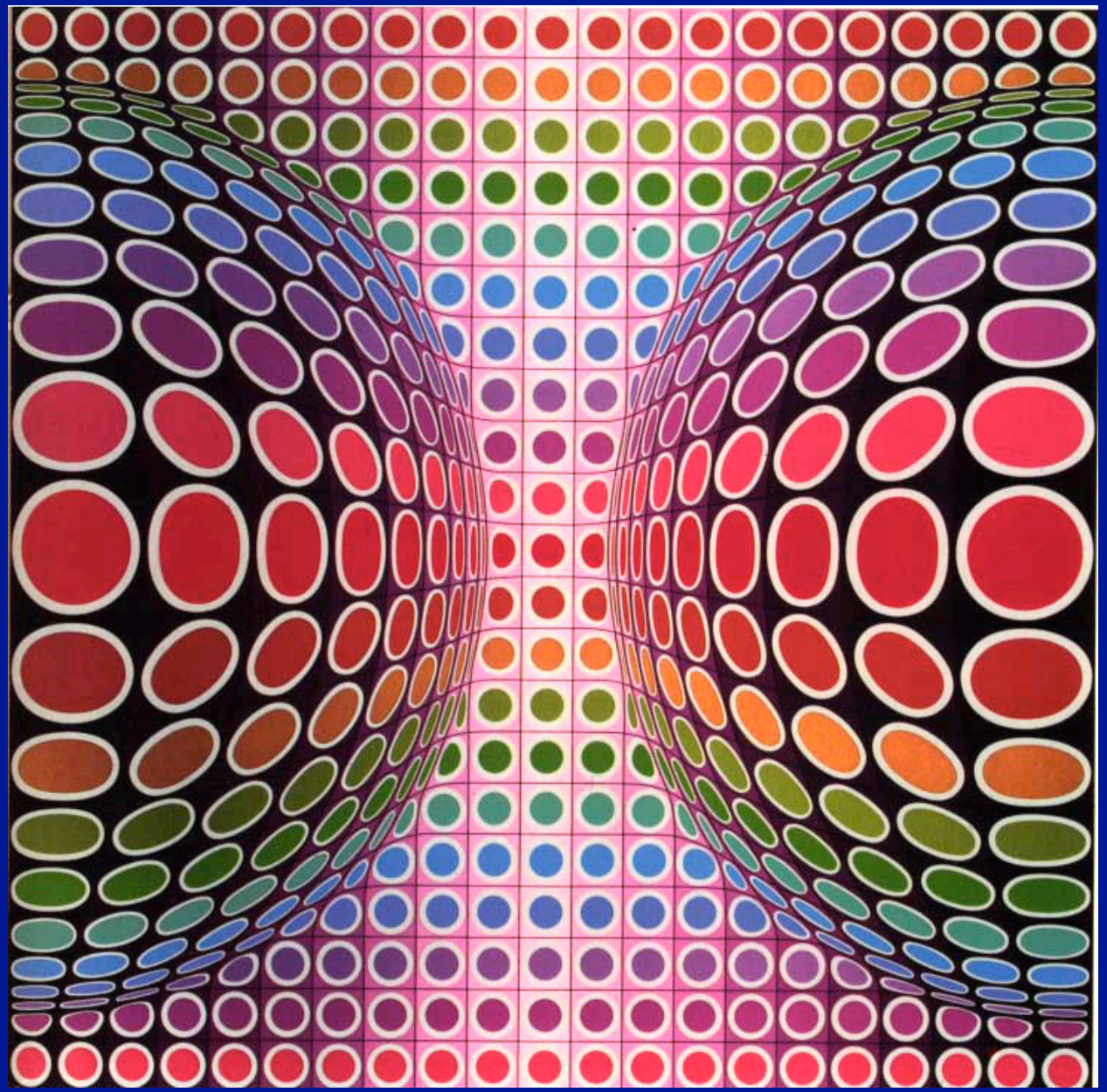

κmin
κmax
S
C

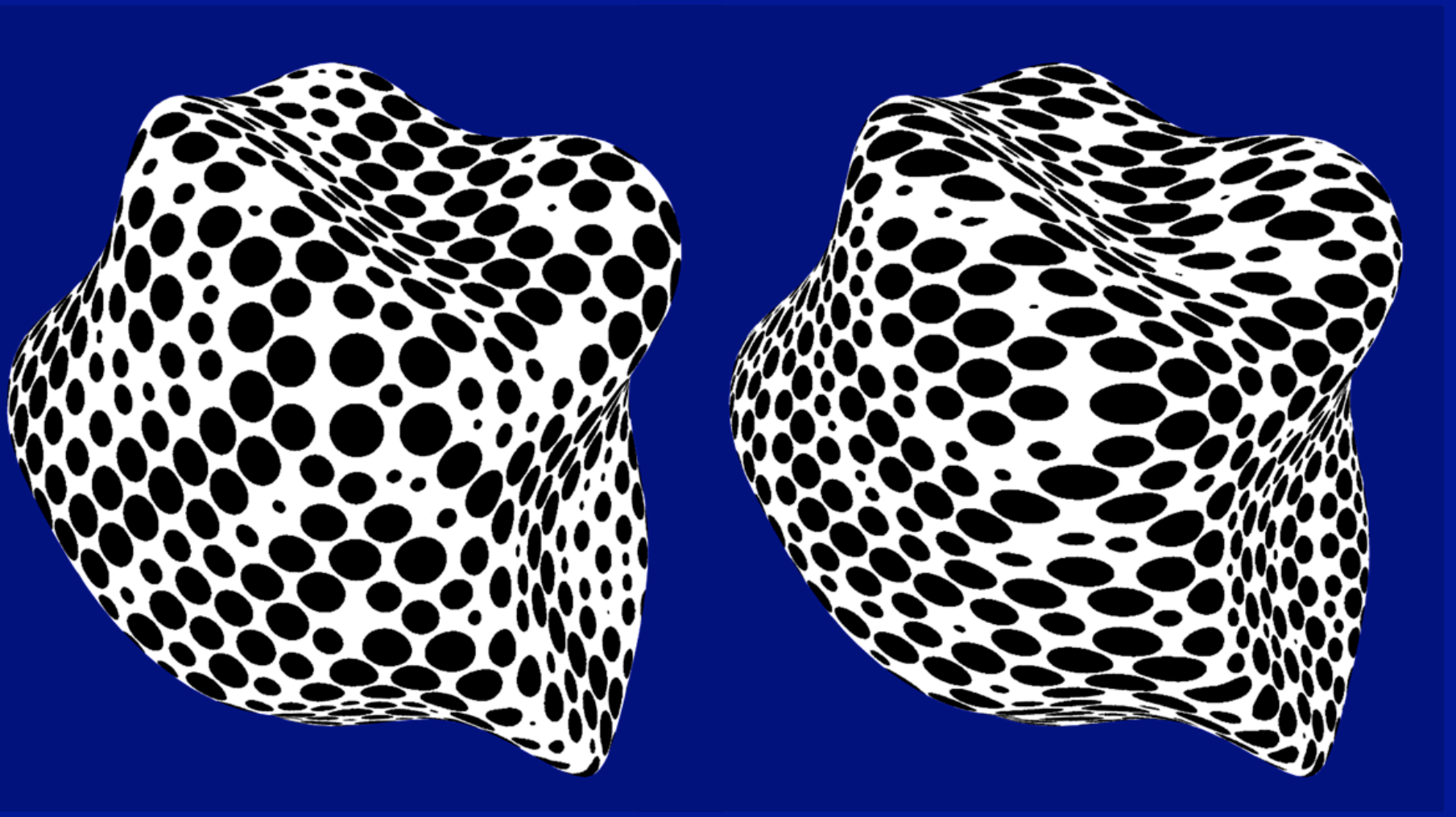

Shape Characteristic
Simulated
Adjusted
3
2
1
-1
-2
-3
-3
-2
-1
1
2
3

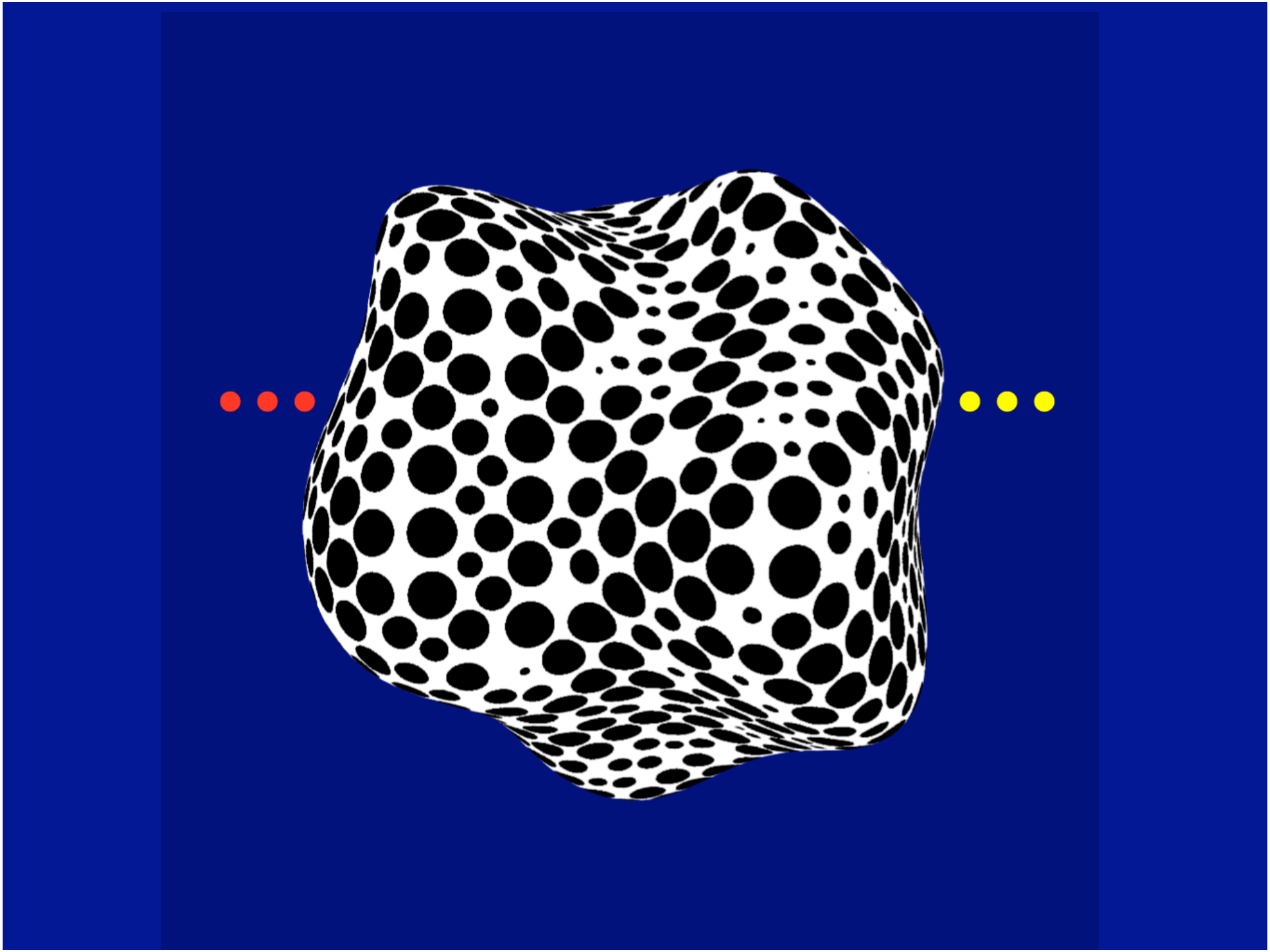

Curvedness
Adjusted
2.5
2
1.5
1
0.5
0.5
0.75
1
1.25
1.5
Simulated

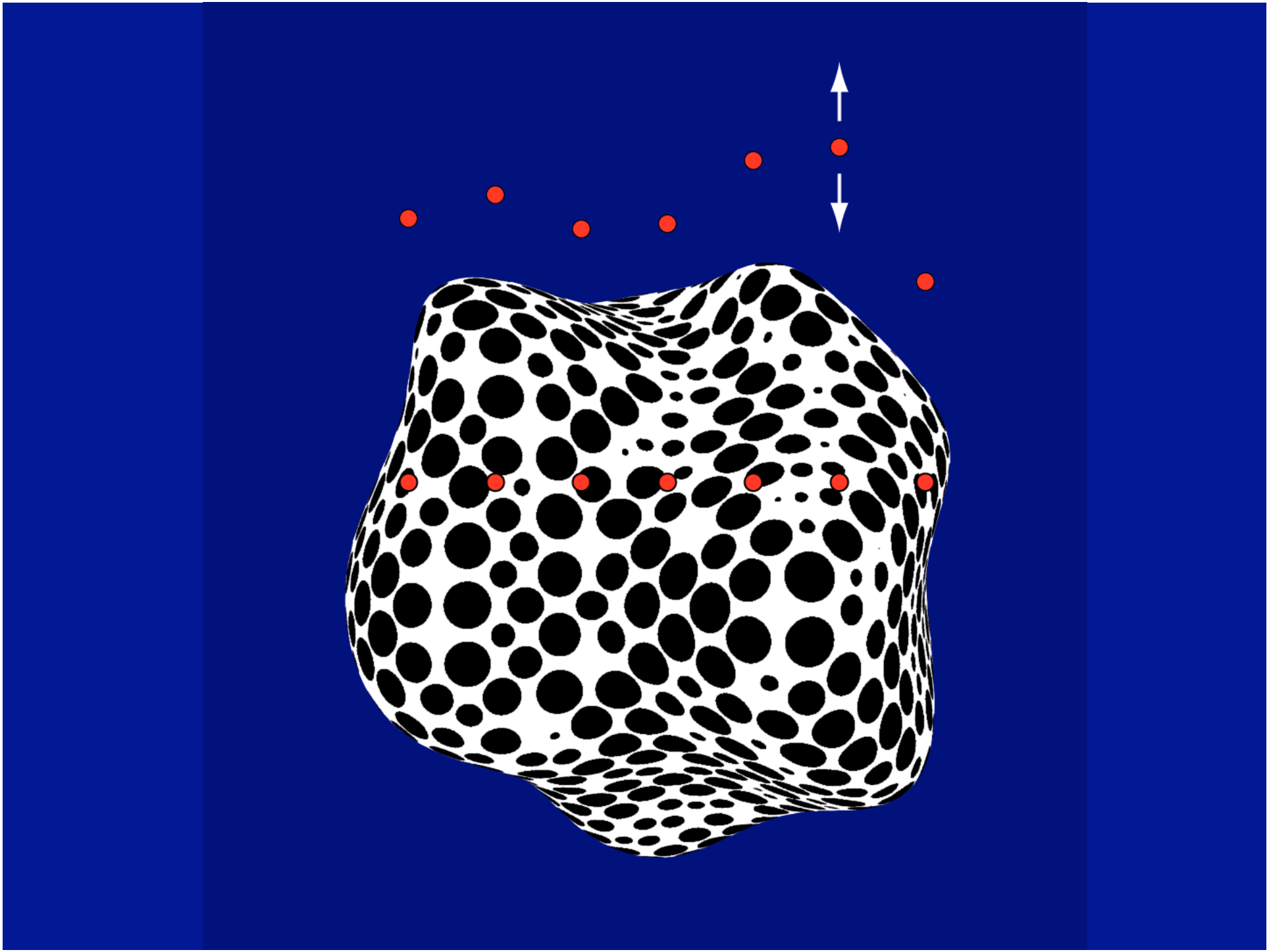

Shape Characteristic
Curvedness
1.0
0.8
0.6
0.4
0.2
0.0
JT
VP
FP
Mean
Observers

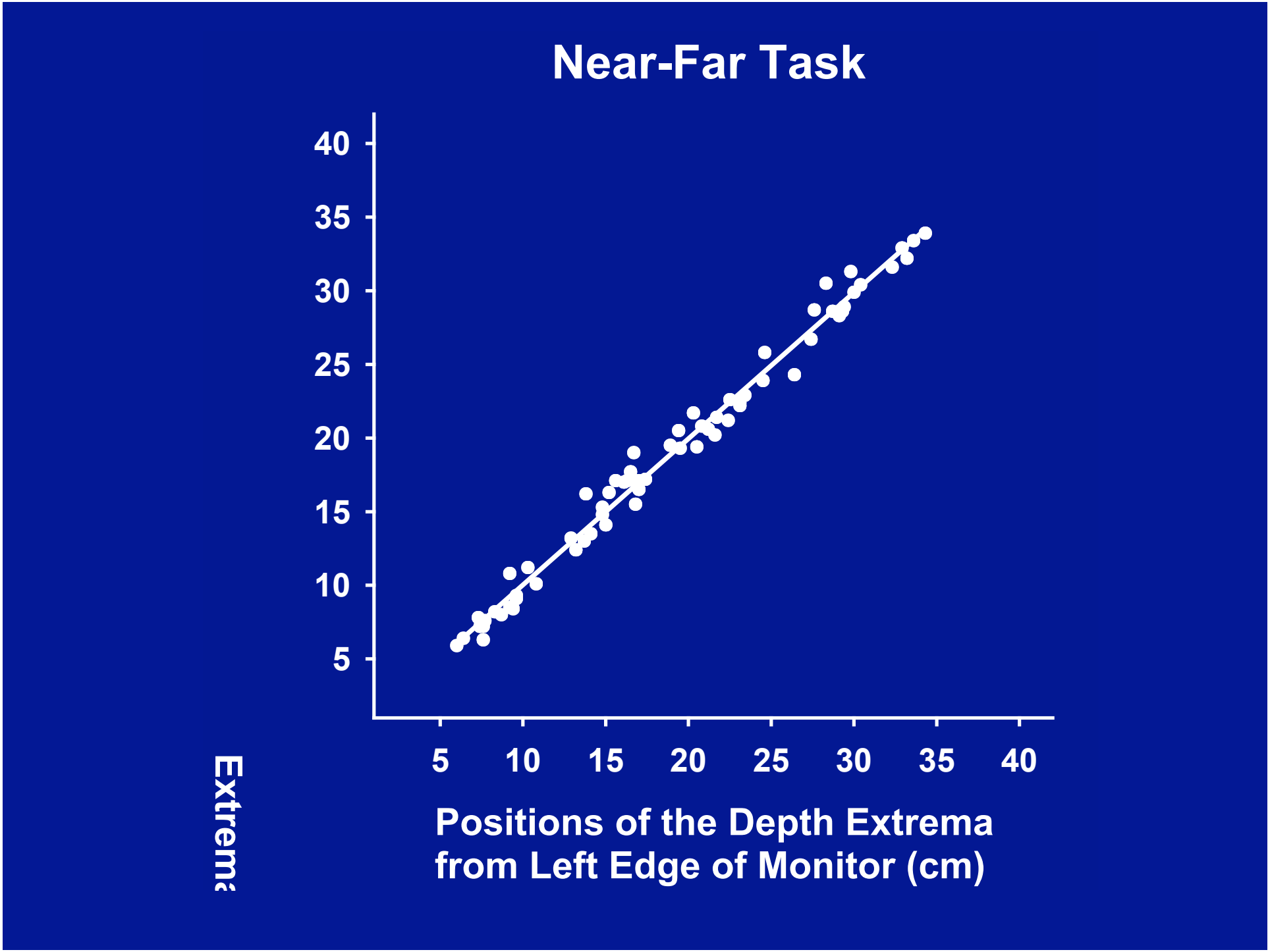
Shape from Texture

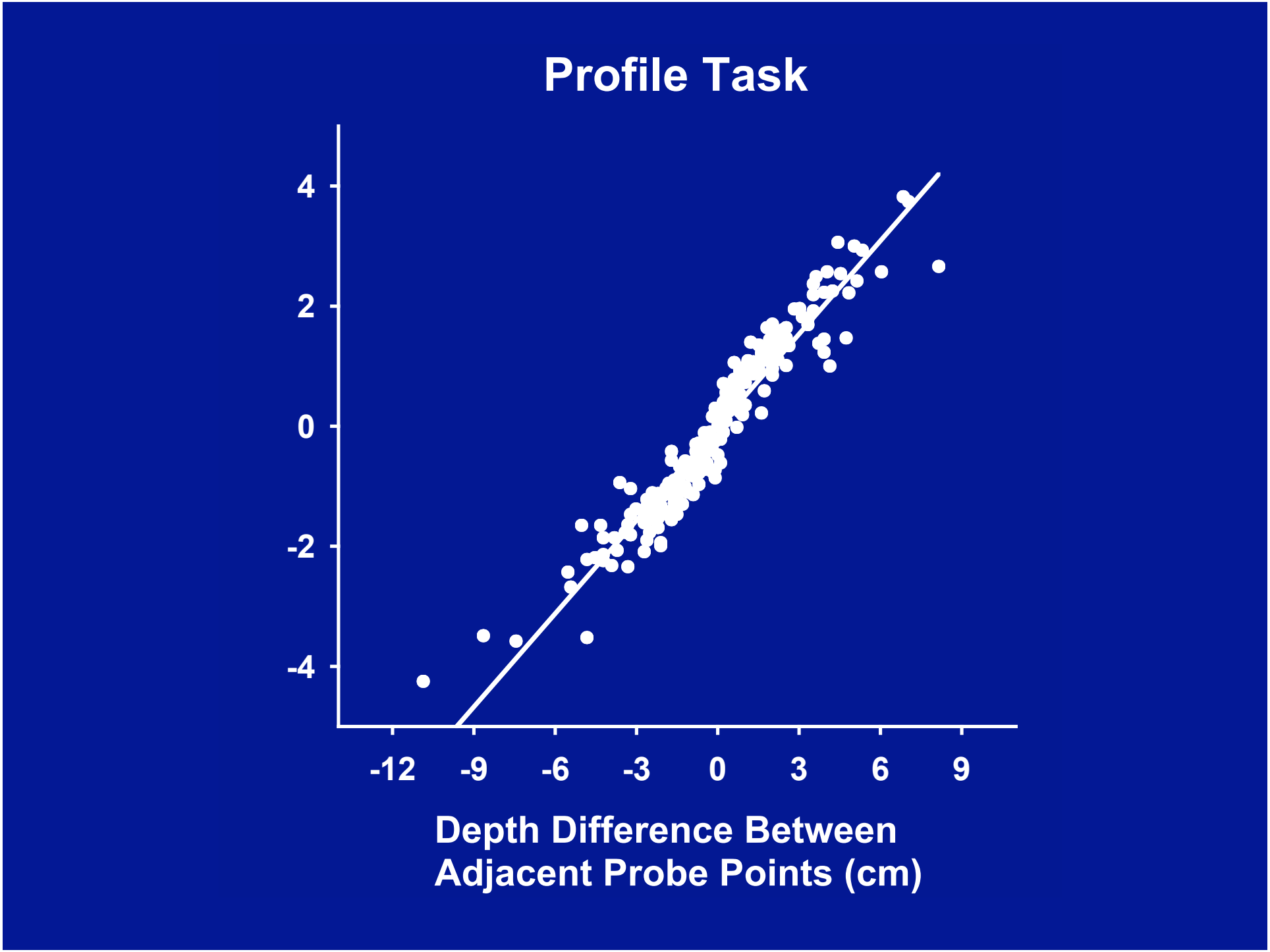


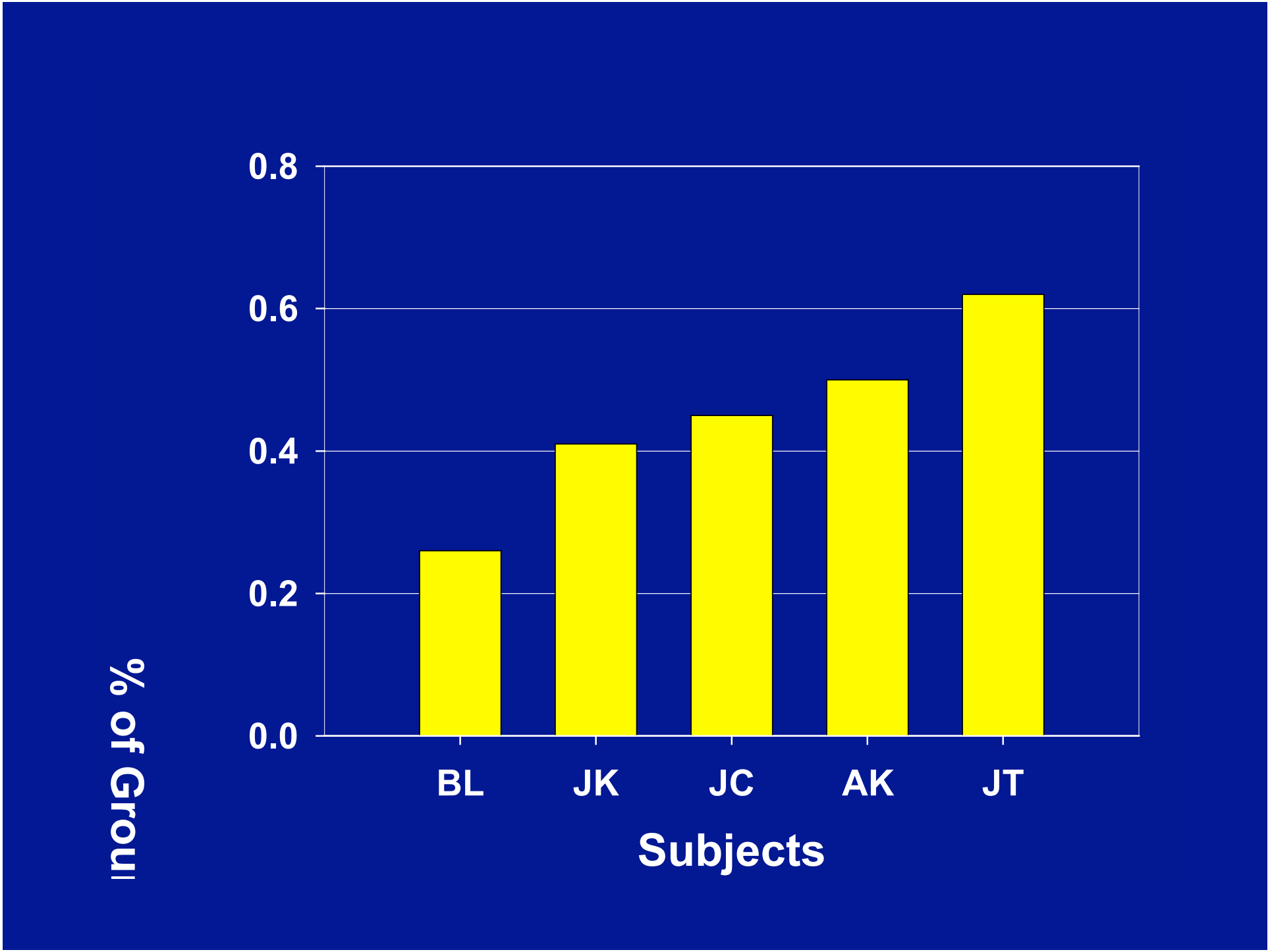







Near-Far Task
40
35
30
25
20
15
10
5
5
10
15
20
25
30
35
40
Positions of the Depth Extrema
from Left Edge of Monitor (cm)
Extrem


Profile Task
4
2
0
-2
-4
-12
-9
-6
-3
0
3
6
9
Depth Difference Between
Adjacent Probe Points (cm)


0.8
0.6
0.4
0.2
0.0
% of Grou
BL
JK
JC
AK
JT
Subjects

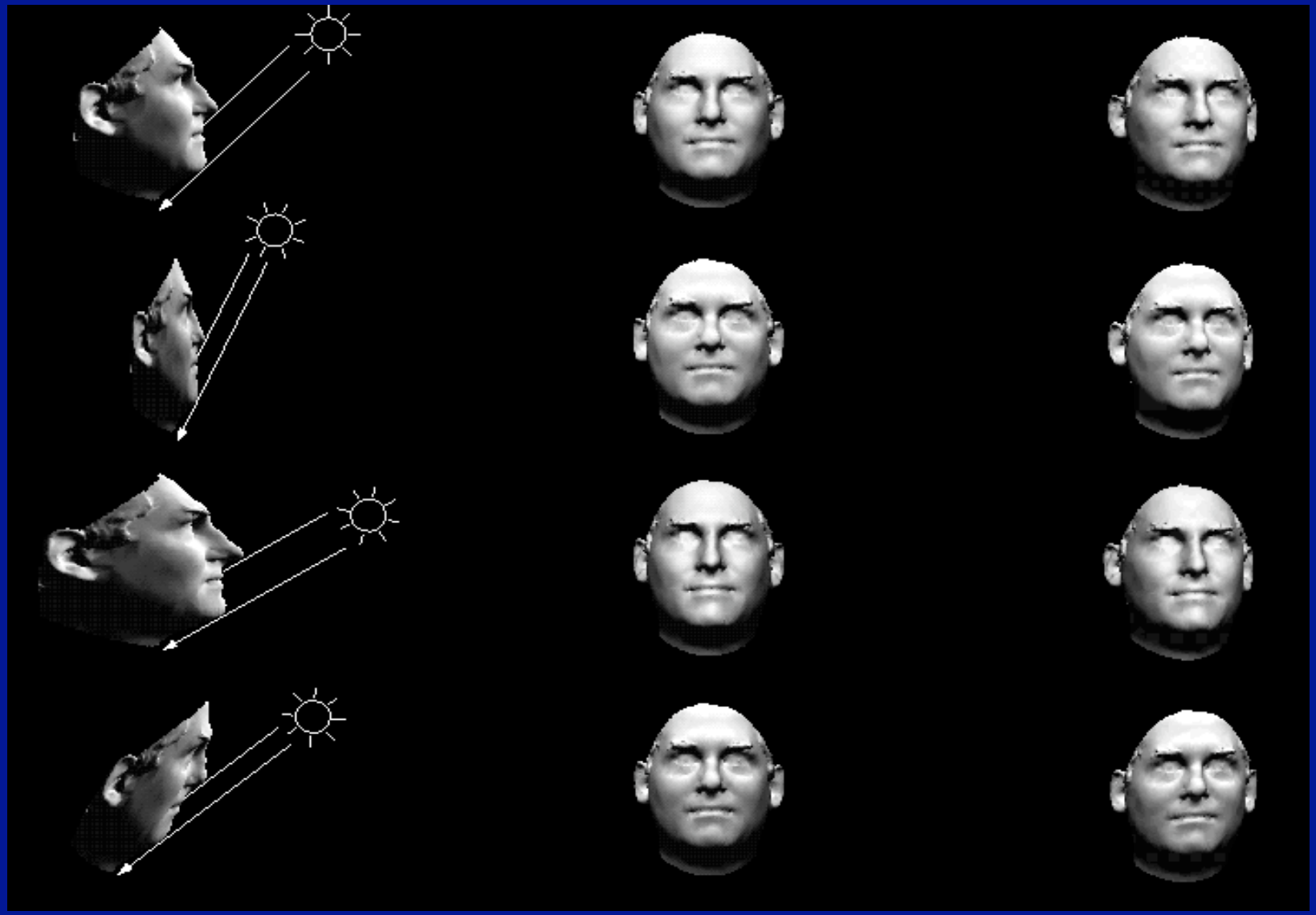
Shape from Shading

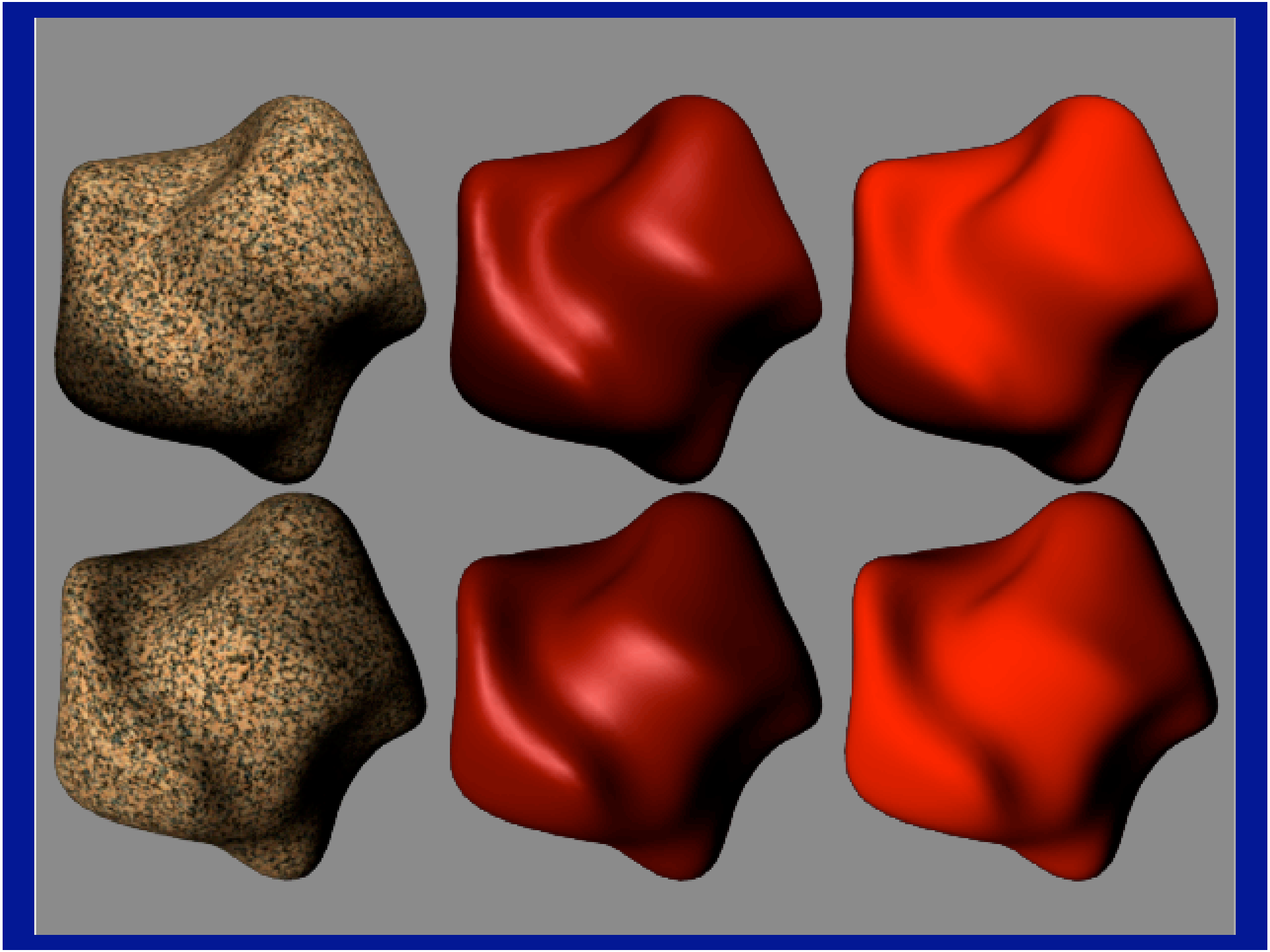
Bas-Relief Ambiguity

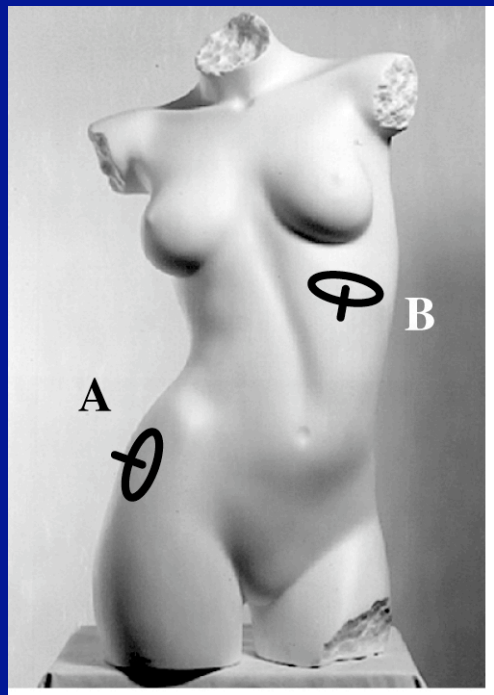
Proposition: For a given pattern of image shading there is an infinite number possible surface interpretations that are all related by an affine transformation.

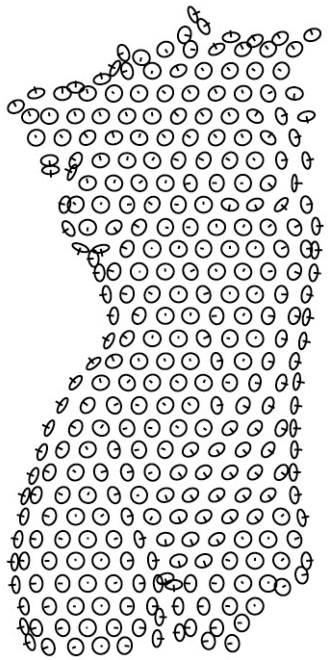
Belhumeur, Kriegman, & Yuille, (1999)

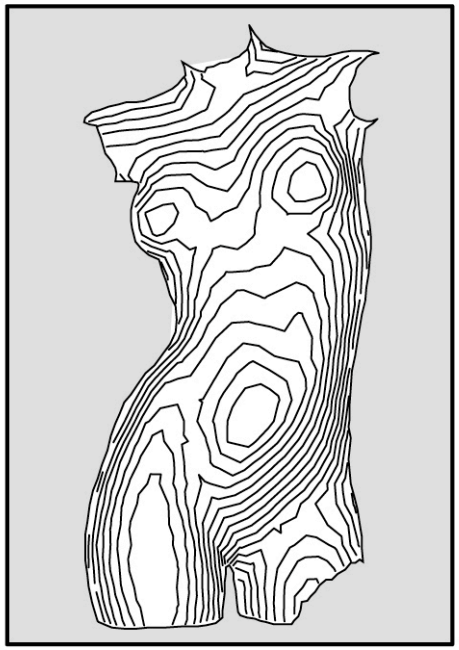
Bas-Relief Ambiguity

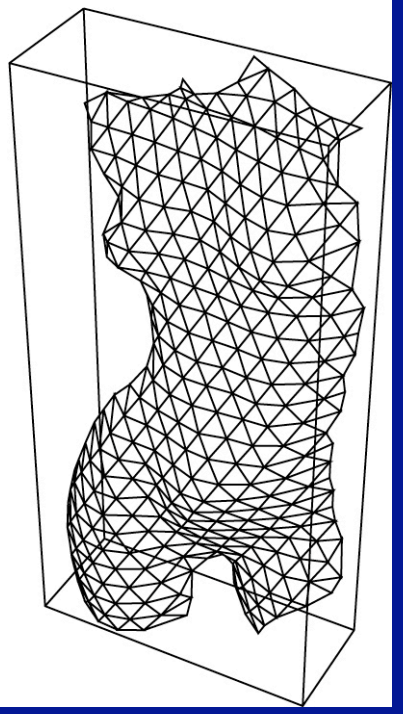


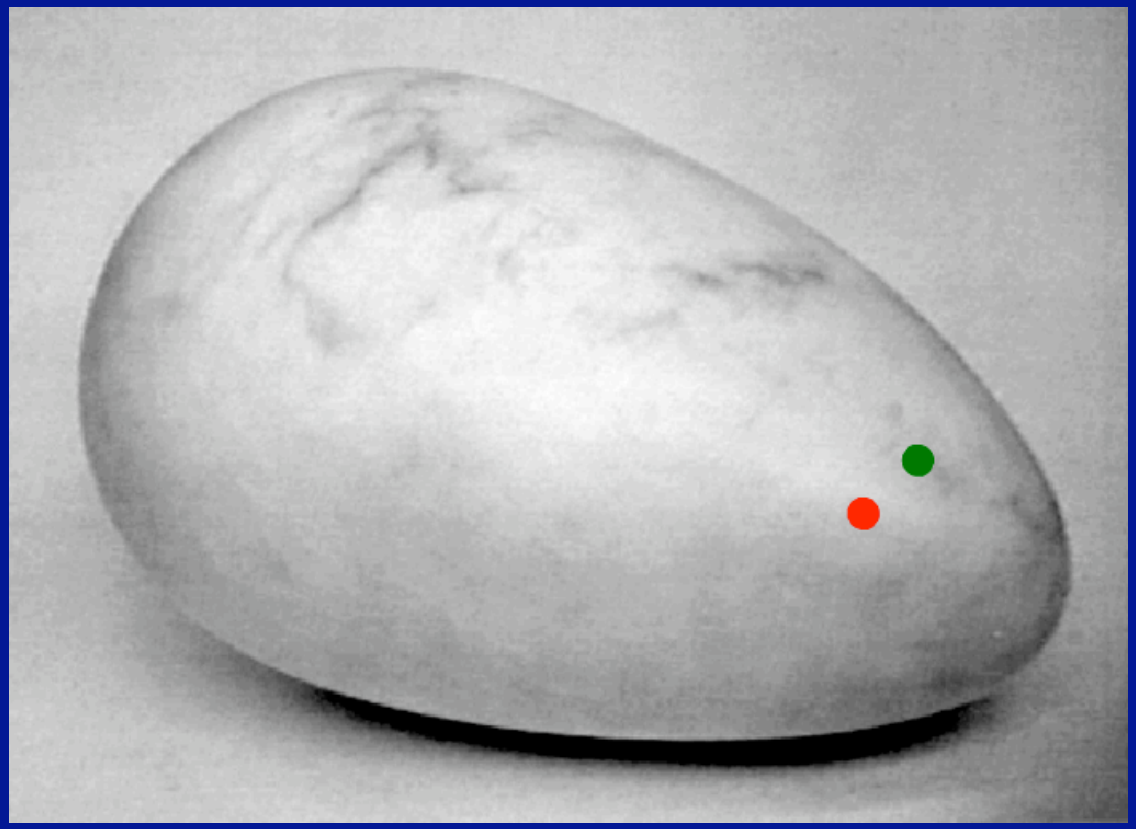


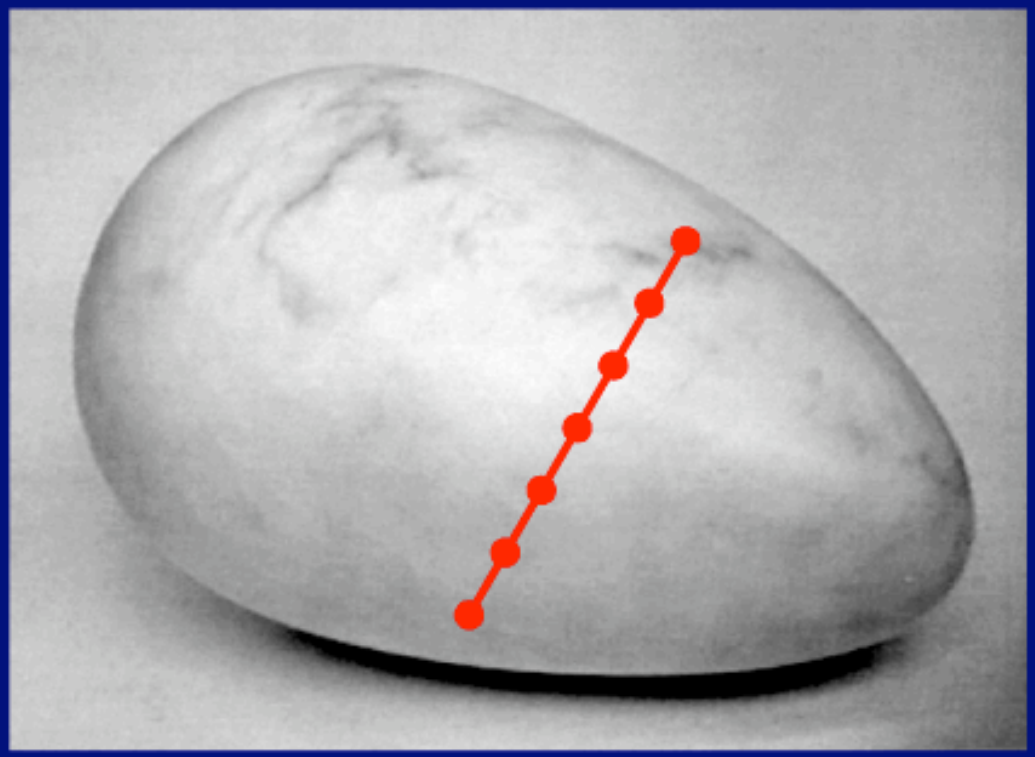

A
B

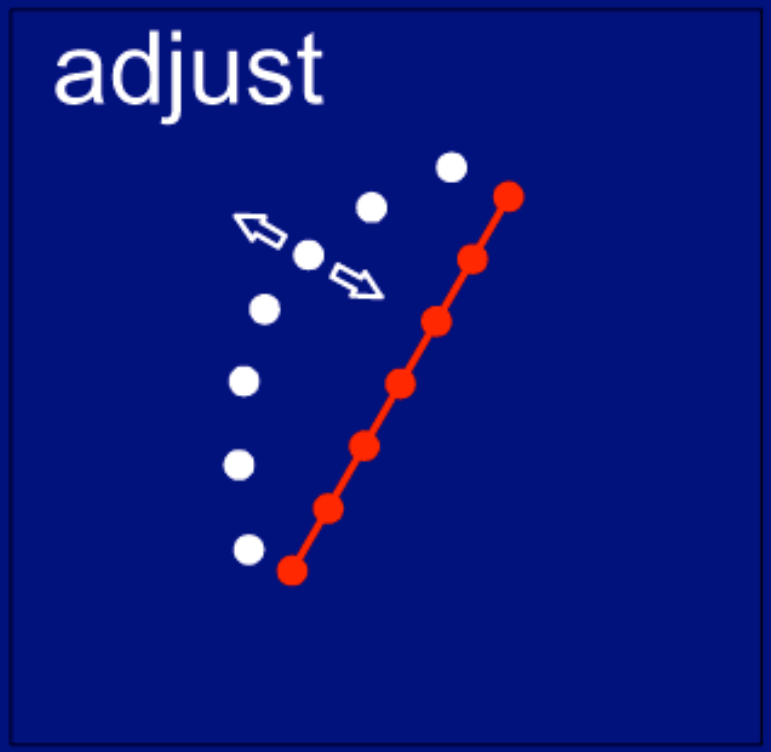


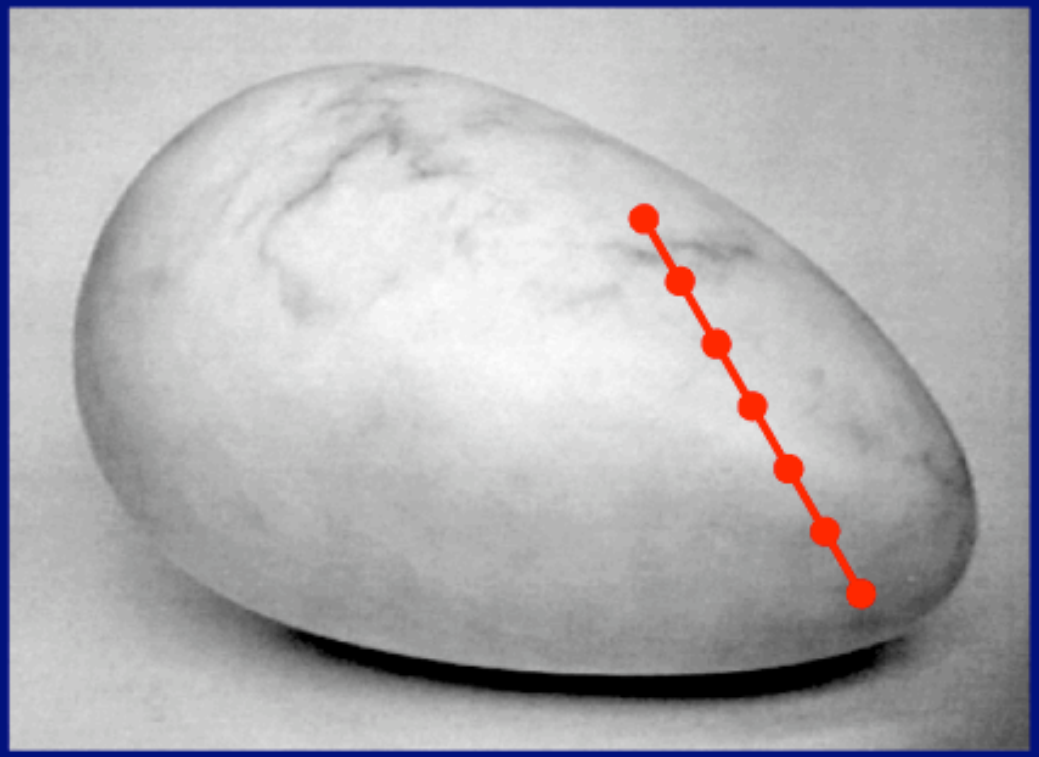




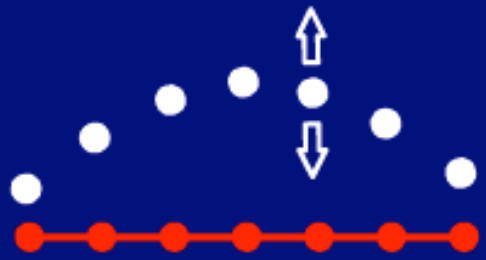




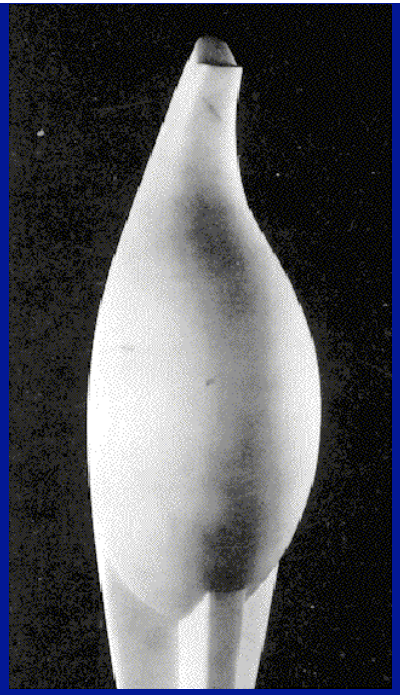

adjust



adjust

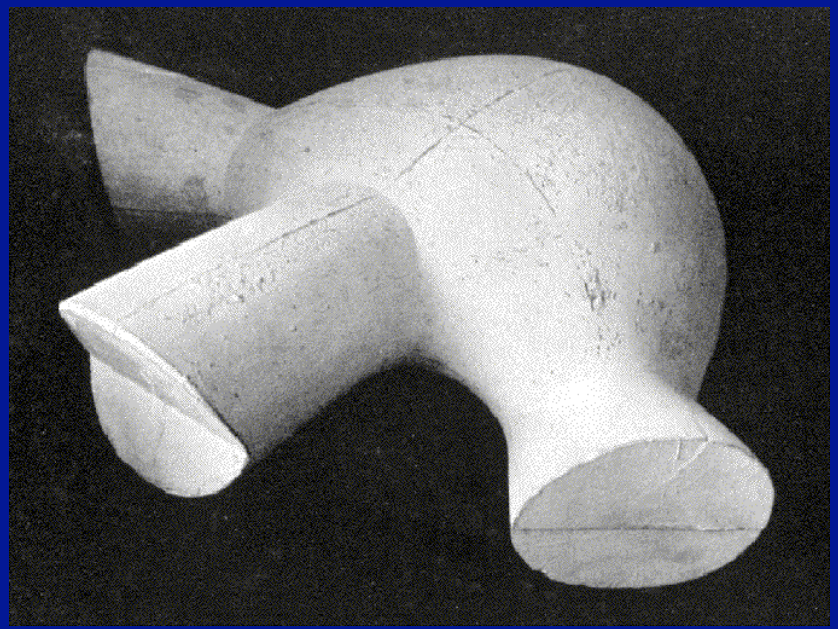


I

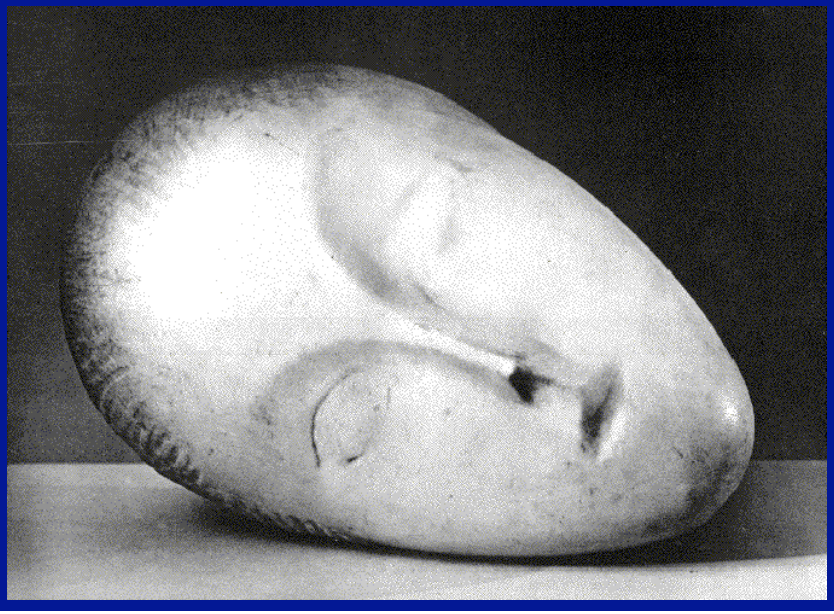


II

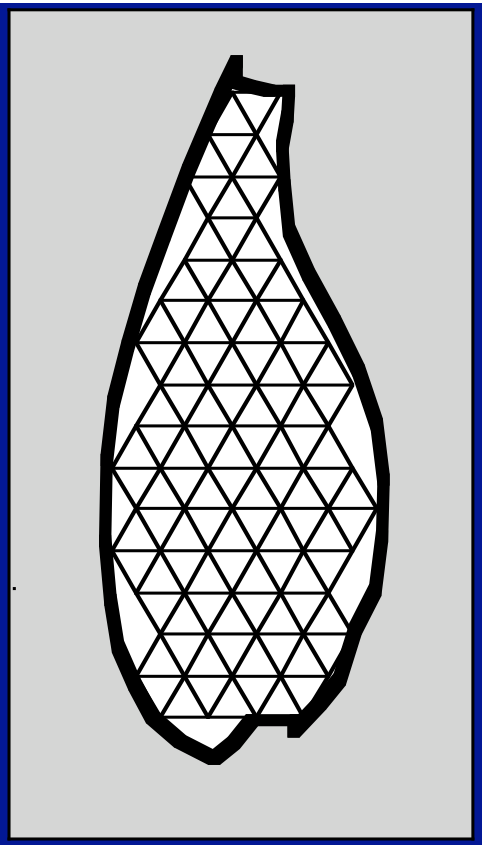
III



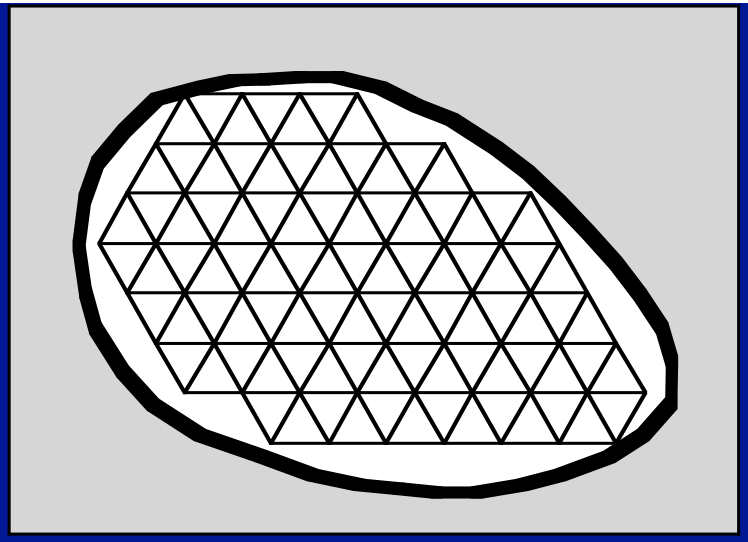
IV



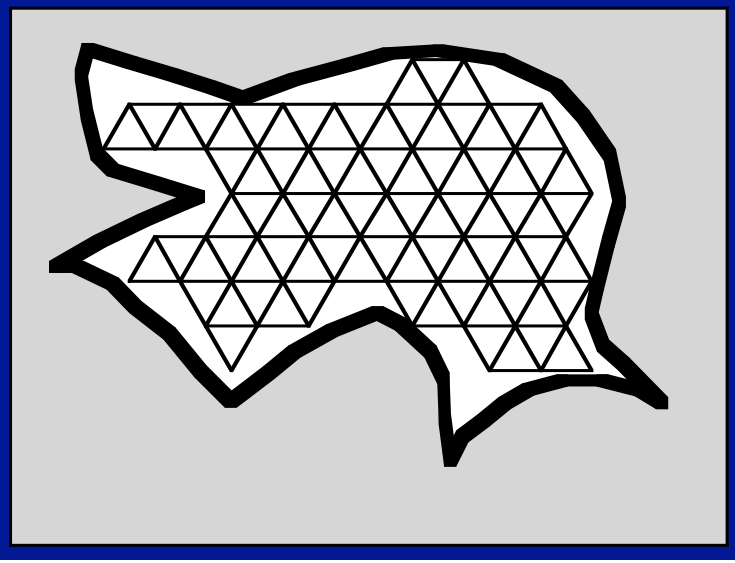
I



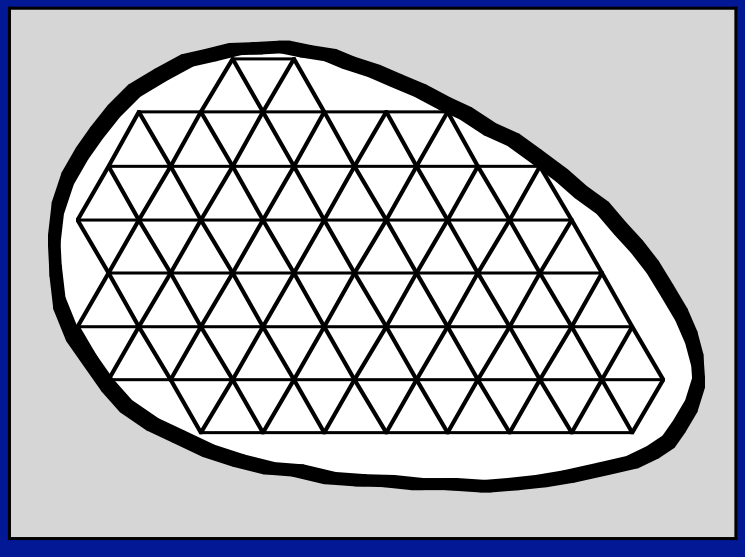
II

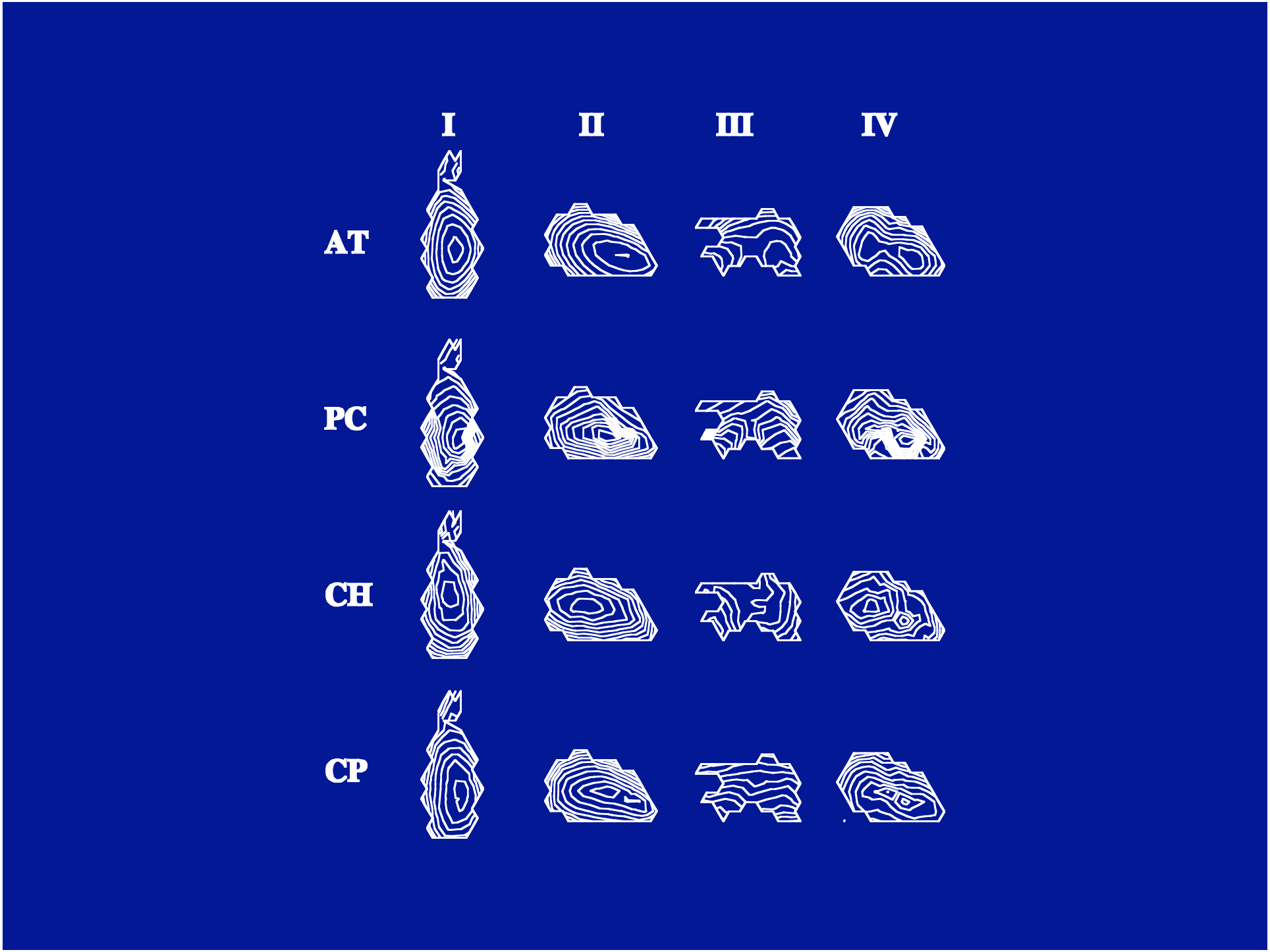


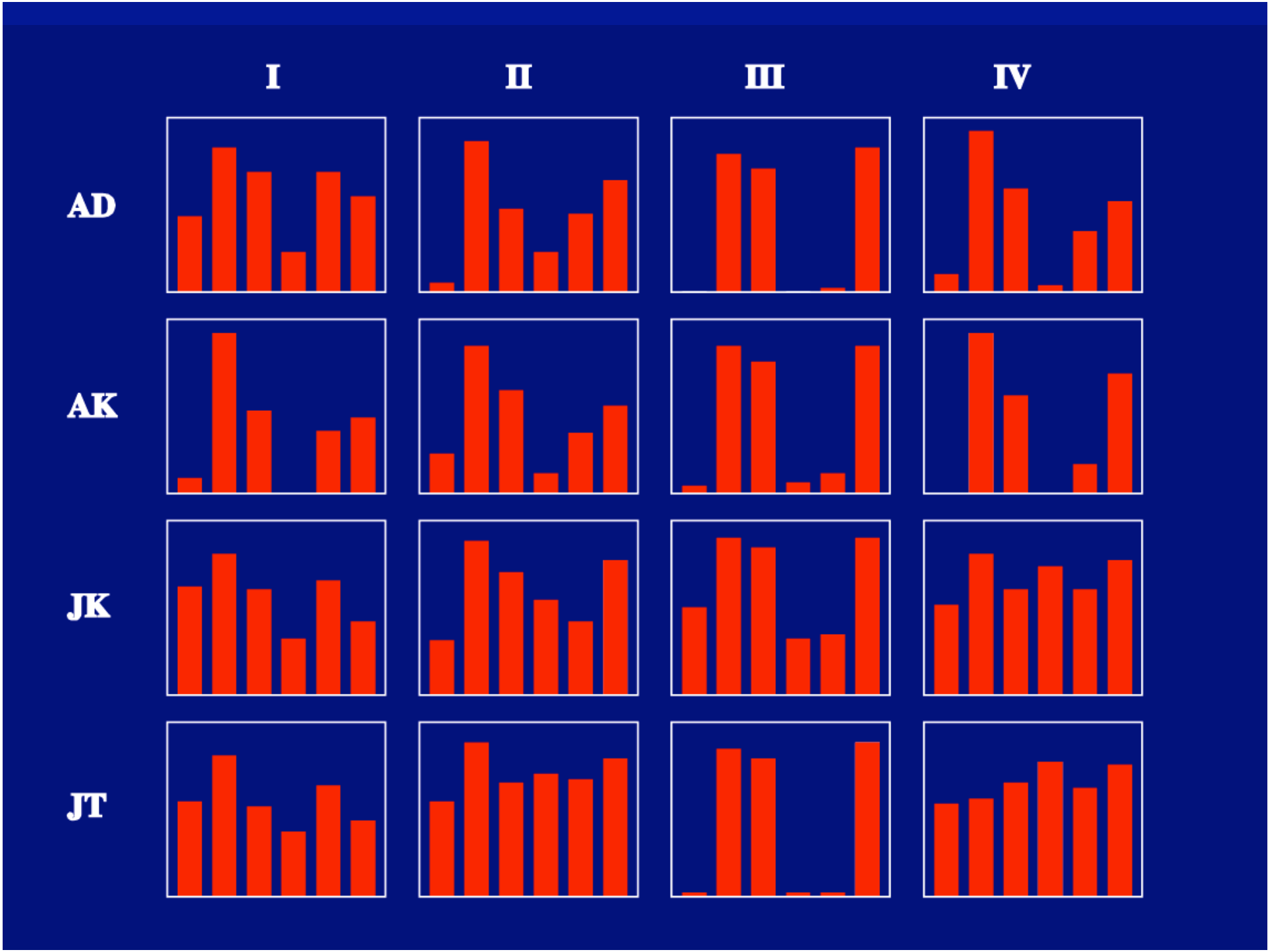
III



IV

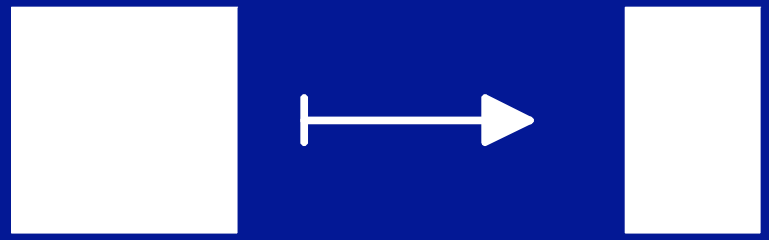



I
II
III
IV
AT
PC
CH
CP

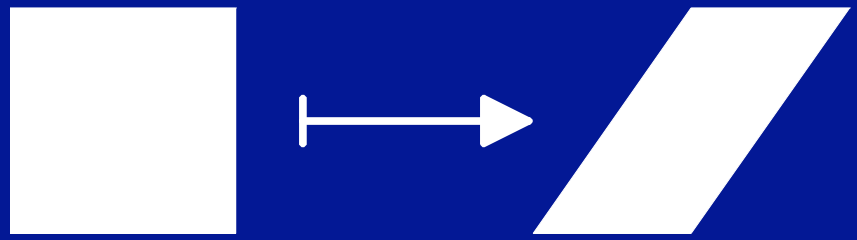

I
II
III
IV
AD
AK
JK
JT

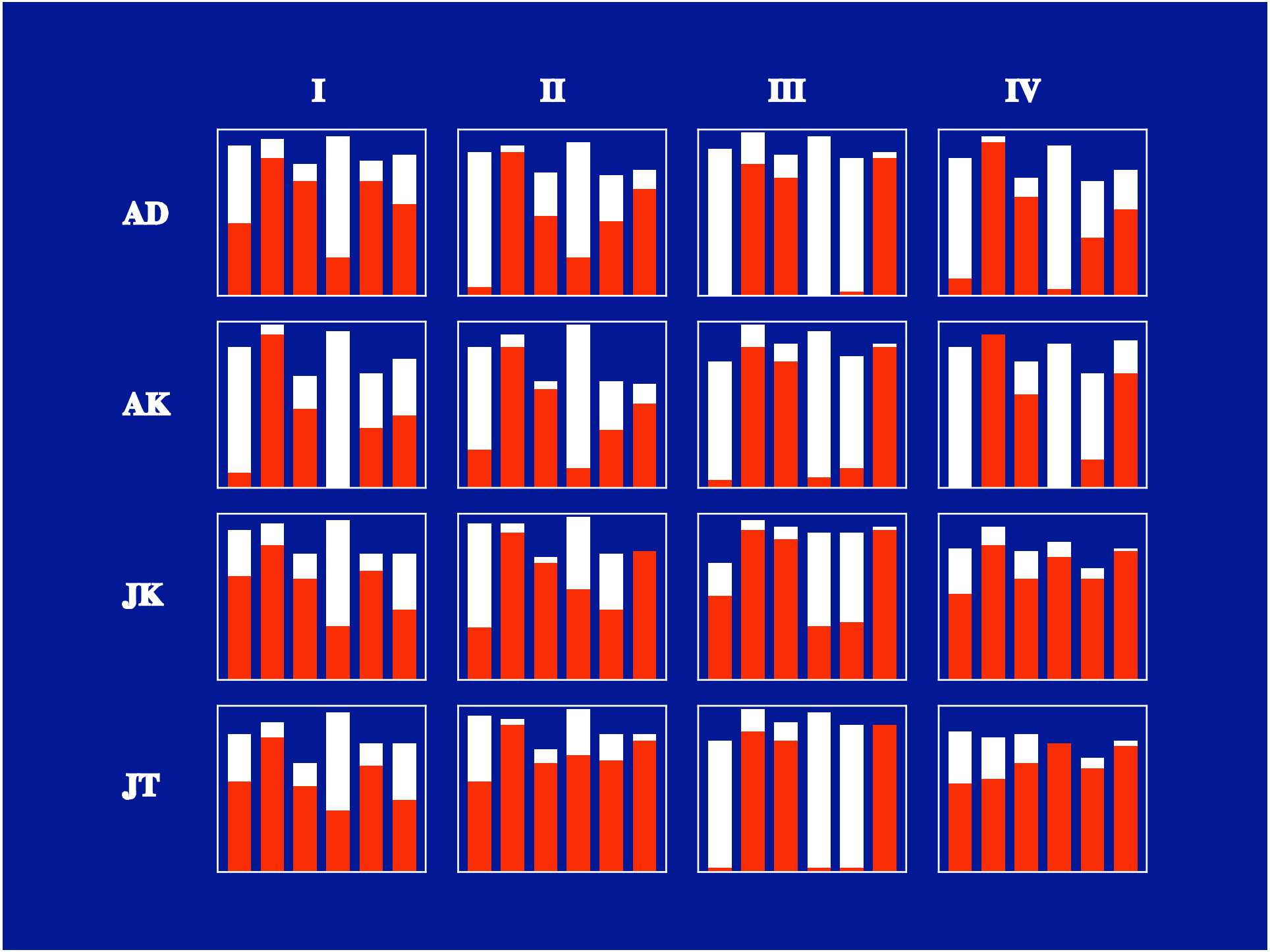
Affine Structure

$$Z' = a\,Z$$

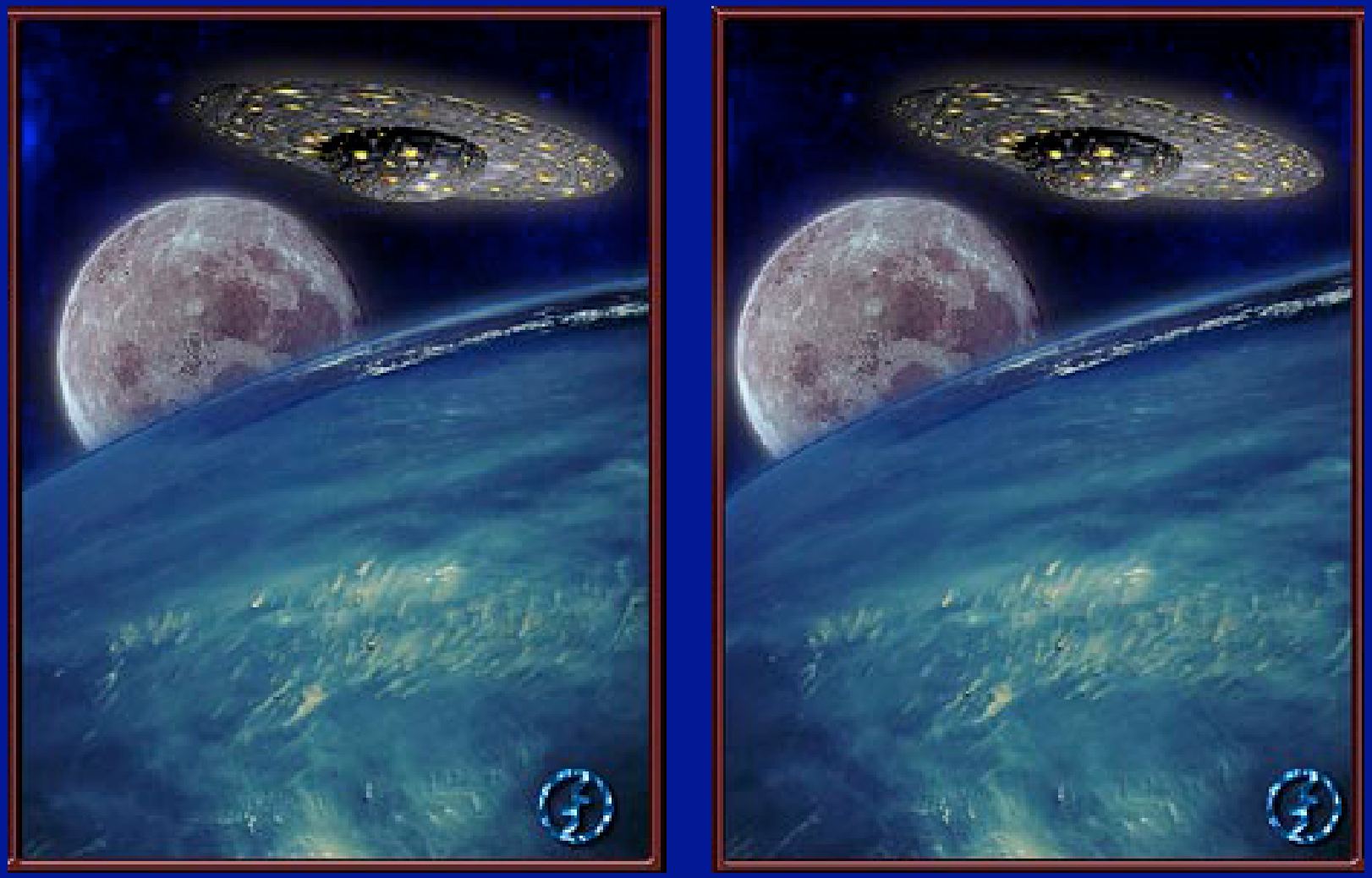


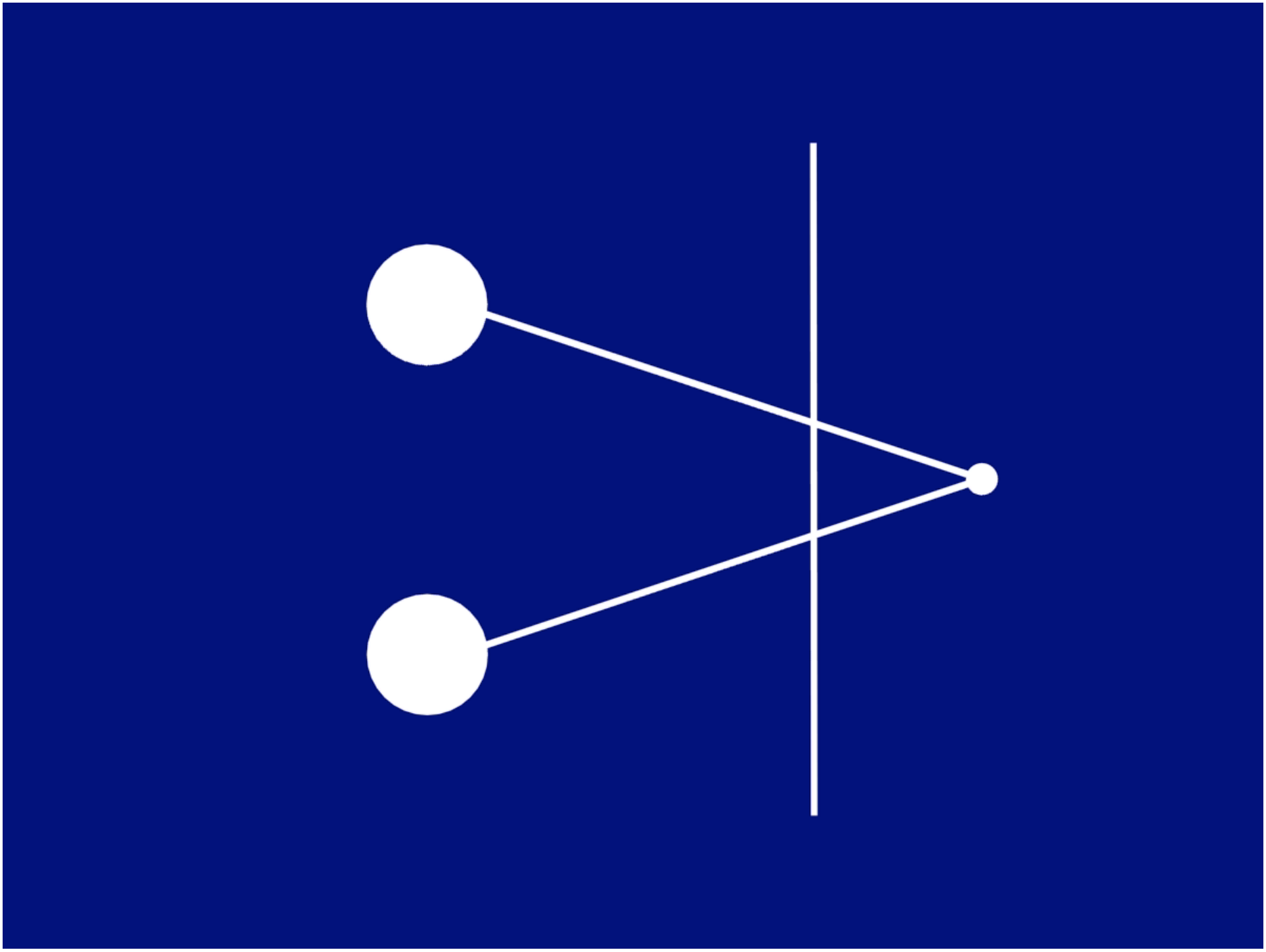
$$Z' = a\,X + b\,Y + c\,Z$$

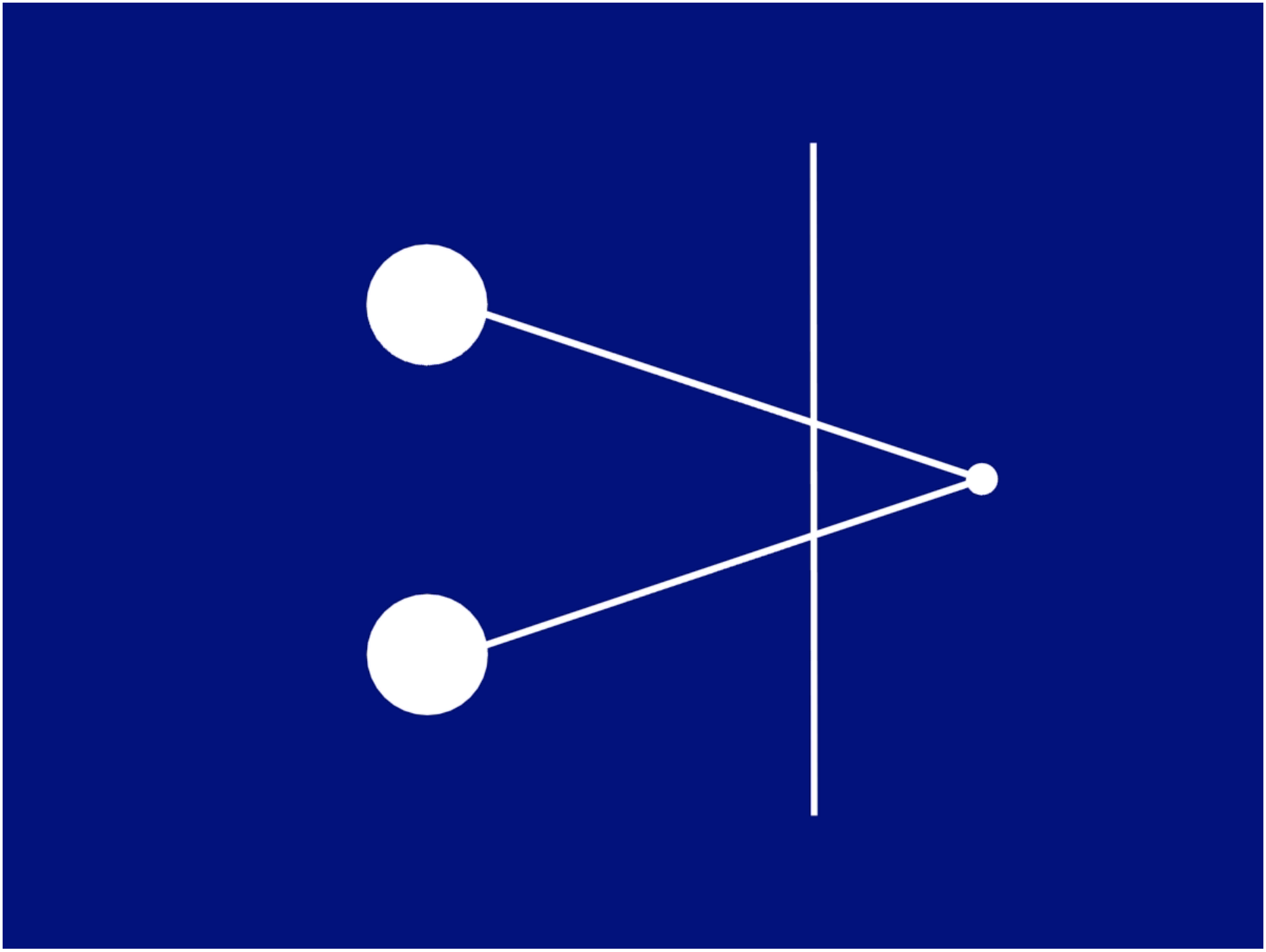


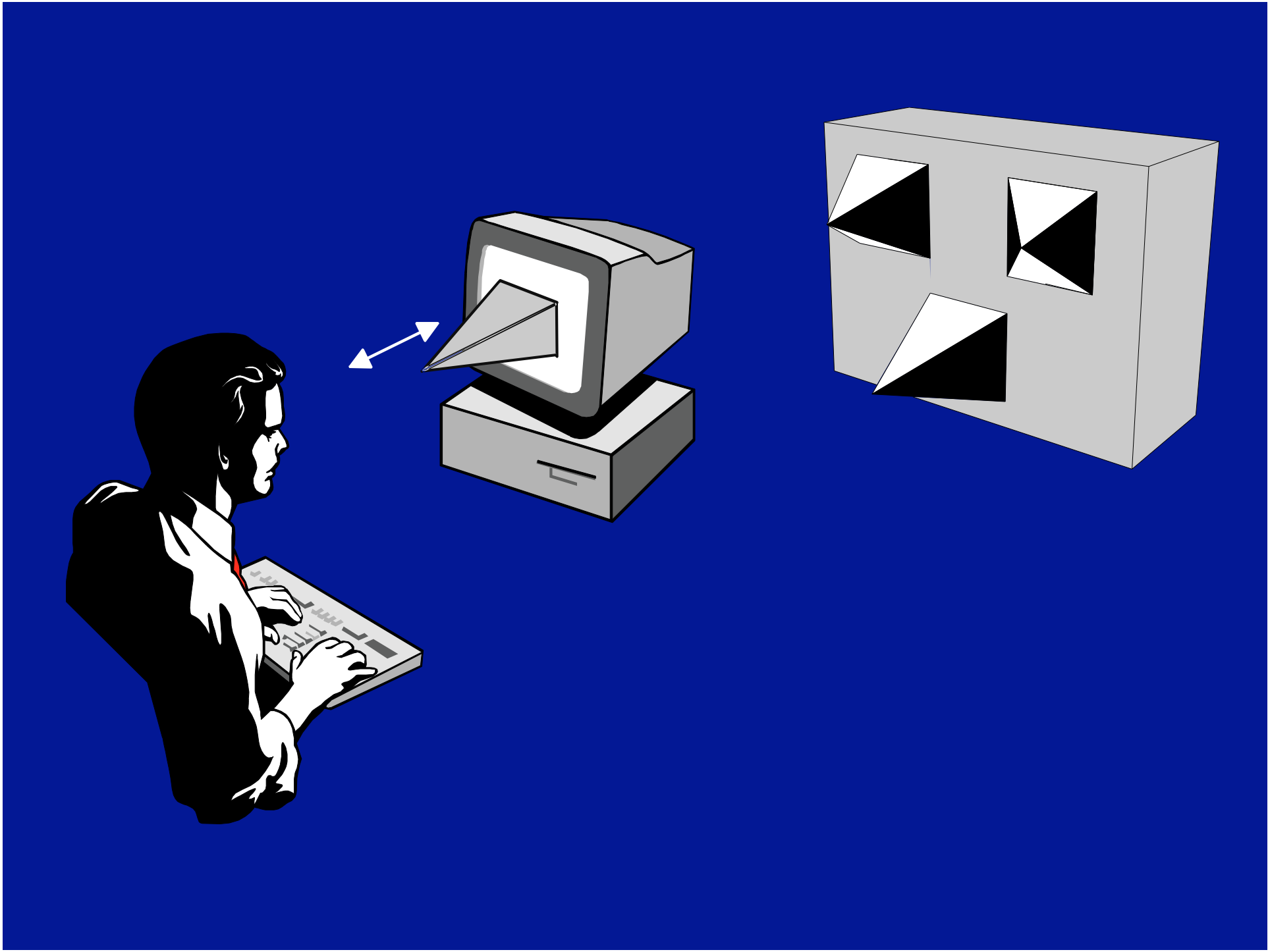

I
II
III
IV
AD
AK
JK
JT

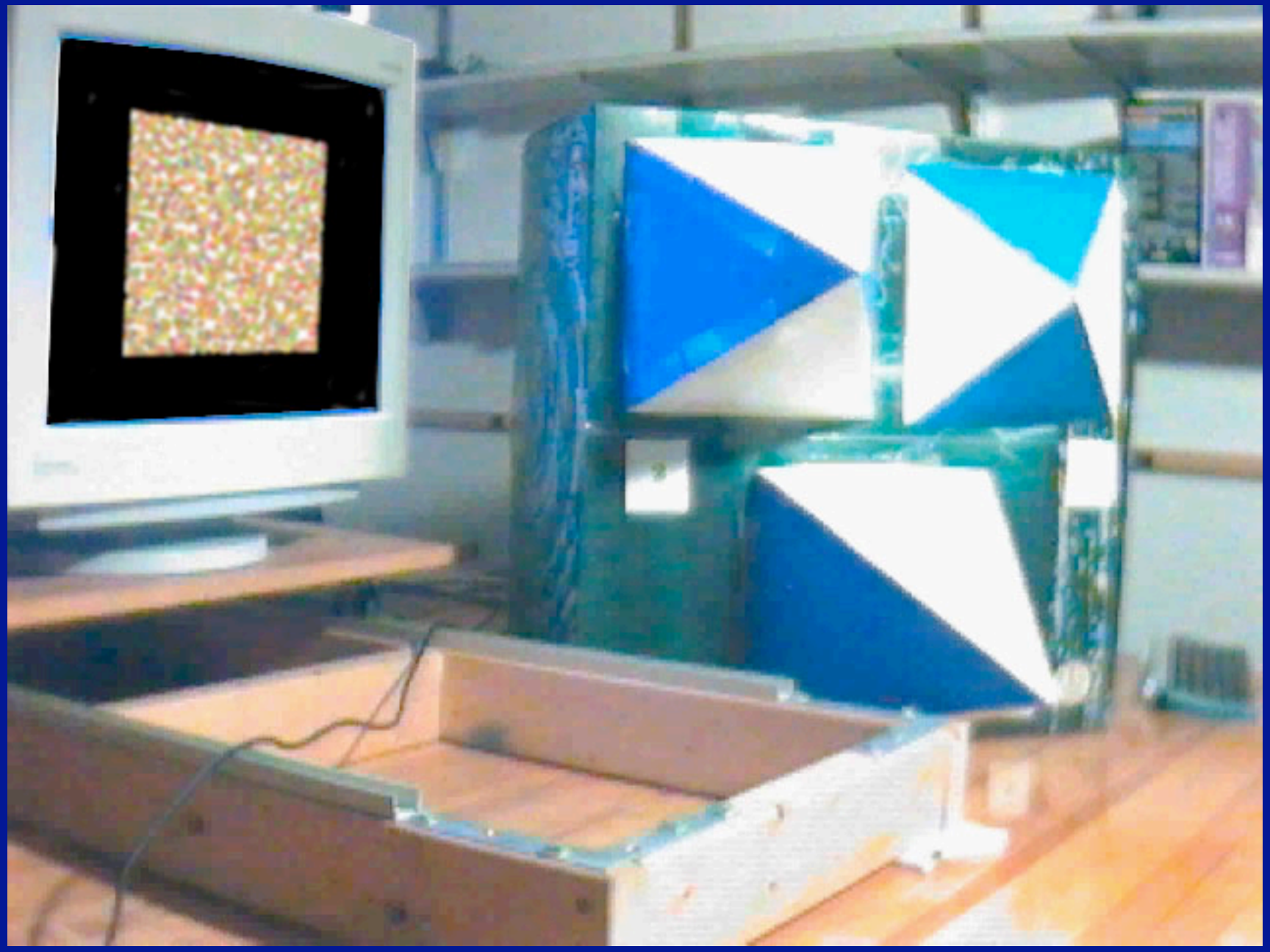
Shape from Binocular Disparity

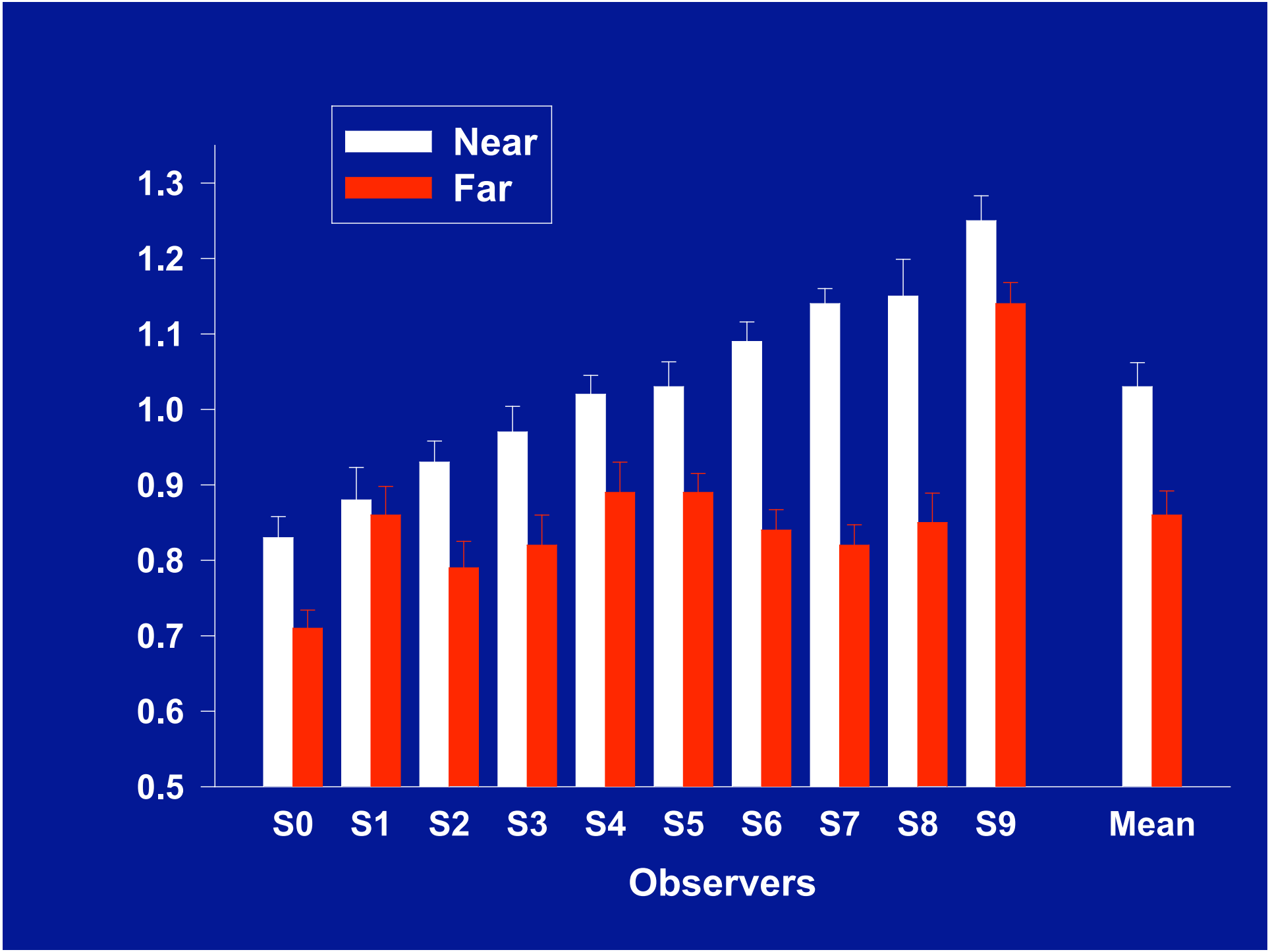


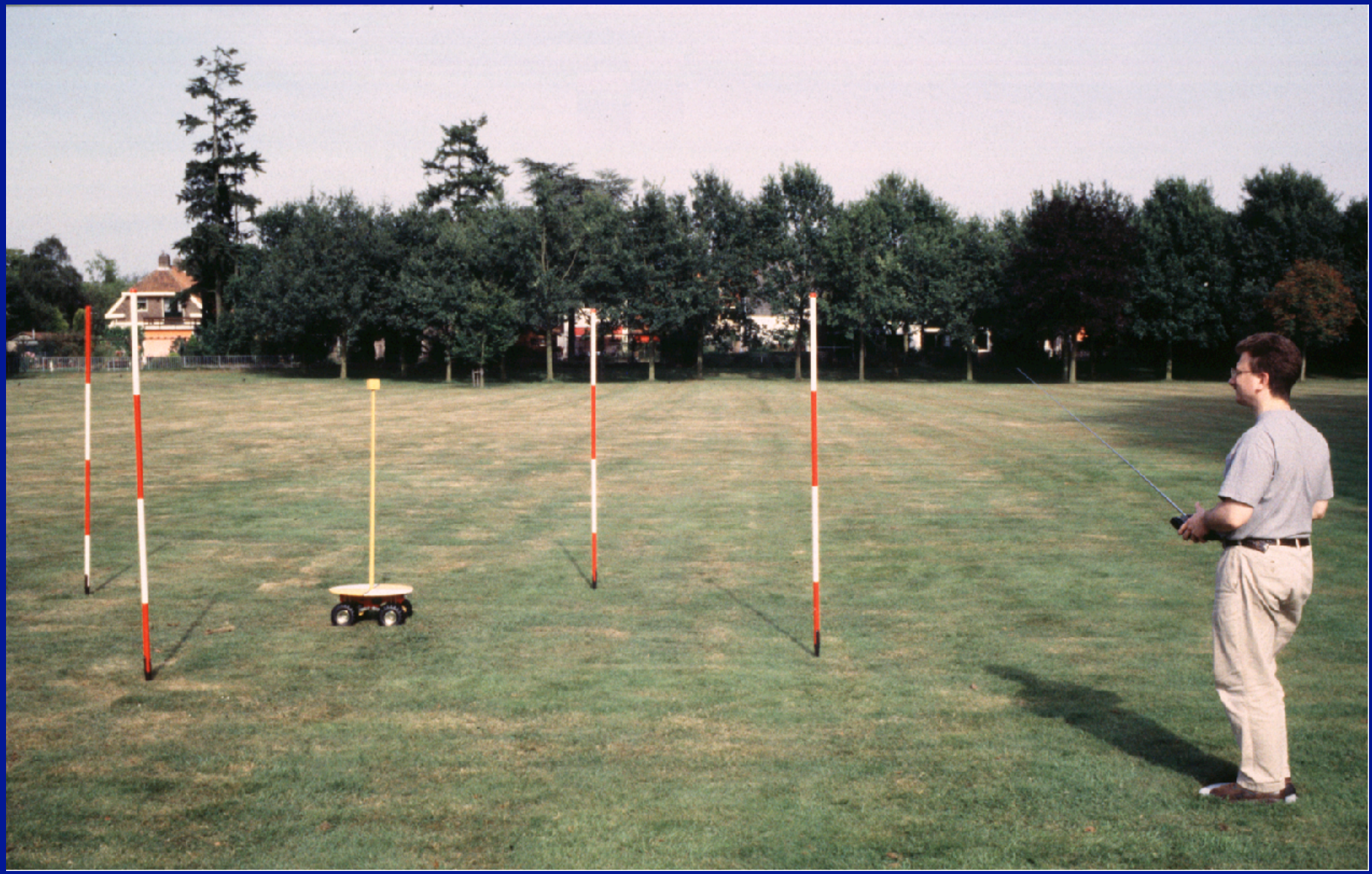




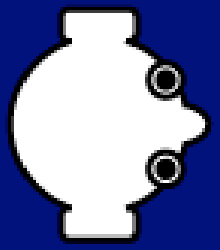


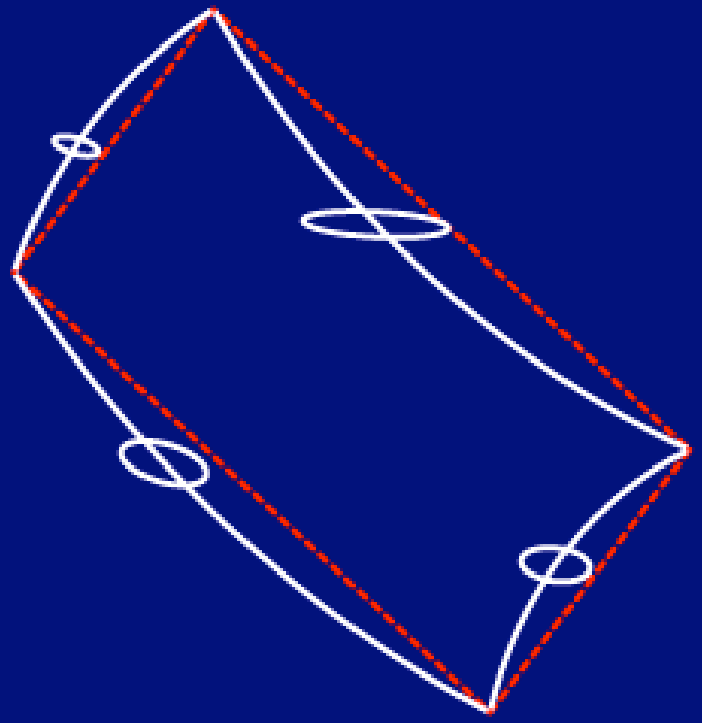



Near
Far
1.3
1.2
1.1
1.0
0.9
0.8
0.7
0.6
0.5
S0
S1
S2
S3
S4
S5
S6
S7
S8
S9
Mean
Observers




TOPCON





Two Possible Conclusions

1) Our perceptual representations are based primarily on nonmetric properties such as affine, ordinal or topological relations.

2) Our perceptual representations have explicit information about 3D metric structure, but that information is inaccurate.