Perception: Psychophysics and Modeling

08 | Object recognition II

Felix Wichmann

Neural Information Processing Group
Eberhard Karls Universität Tübingen

Overview

The Problems of Perceiving and Recognising Objects **(VL07-Object Recognition 1)**

Mid-level vision **(VL07-Object Recognition 1)**

- What are "edges" and (illusionary) "contours"?
- Gestalt psychology and "Gestalt laws" of perceptual organisation

More on *mid-level vision* **(VL08-Object Recognition 2)**

- Accidental viewpoint and non-accidental features
- Figure-ground, occlusion, wholes and parts
- Texture segmentation, grouping and camouflage

Neuroscience of object recognition **(VL08-Object Recognition 2)**

Object representation **(VL09-Object Recognition 3)**

- Structural description models
- View-based models

Object recognition by algorithms: DNNs **(VL10-Object Recognition 4)**

Overview

The Problems of Perceiving and Recognising Objects **(VL07-Object Recognition 1)**

Mid-level vision **(VL07-Object Recognition 1)**

- What are "edges" and (illusionary) "contours"?
- Gestalt psychology and "Gestalt laws" of perceptual organisation

More on *mid-level vision* **(VL08-Object Recognition 2)**

- Accidental viewpoint and non-accidental features
- Figure-ground, occlusion, wholes and parts
- Texture segmentation, grouping and camouflage

Neuroscience of object recognition **(VL08-Object Recognition 2)**

Object representation **(VL09-Object Recognition 3)**

- Structural description models
- View-based models

Object recognition by algorithms: DNNs **(VL10-Object Recognition 4)**

Supplementary Literature

Barlow, H. (1972). Single units and sensation: A neuron doctrine for perceptual psychology. *Perception*, 1: 371-394.

Desimone, R. & Ungerleider, L. G. (1989). Neural mechanisms of visual processing in monkeys. In Boller, F. & Grafman, J., editors, *Handbook of Neurophysiology*, Vol. 2, chapter 14, pages 267-299. Elsevier, Amsterdam.

Douglas, R. J. and Martin, K. A. C. (1991). Opening the grey box. *Trends in Neurosciences*, 14(7):286–293.

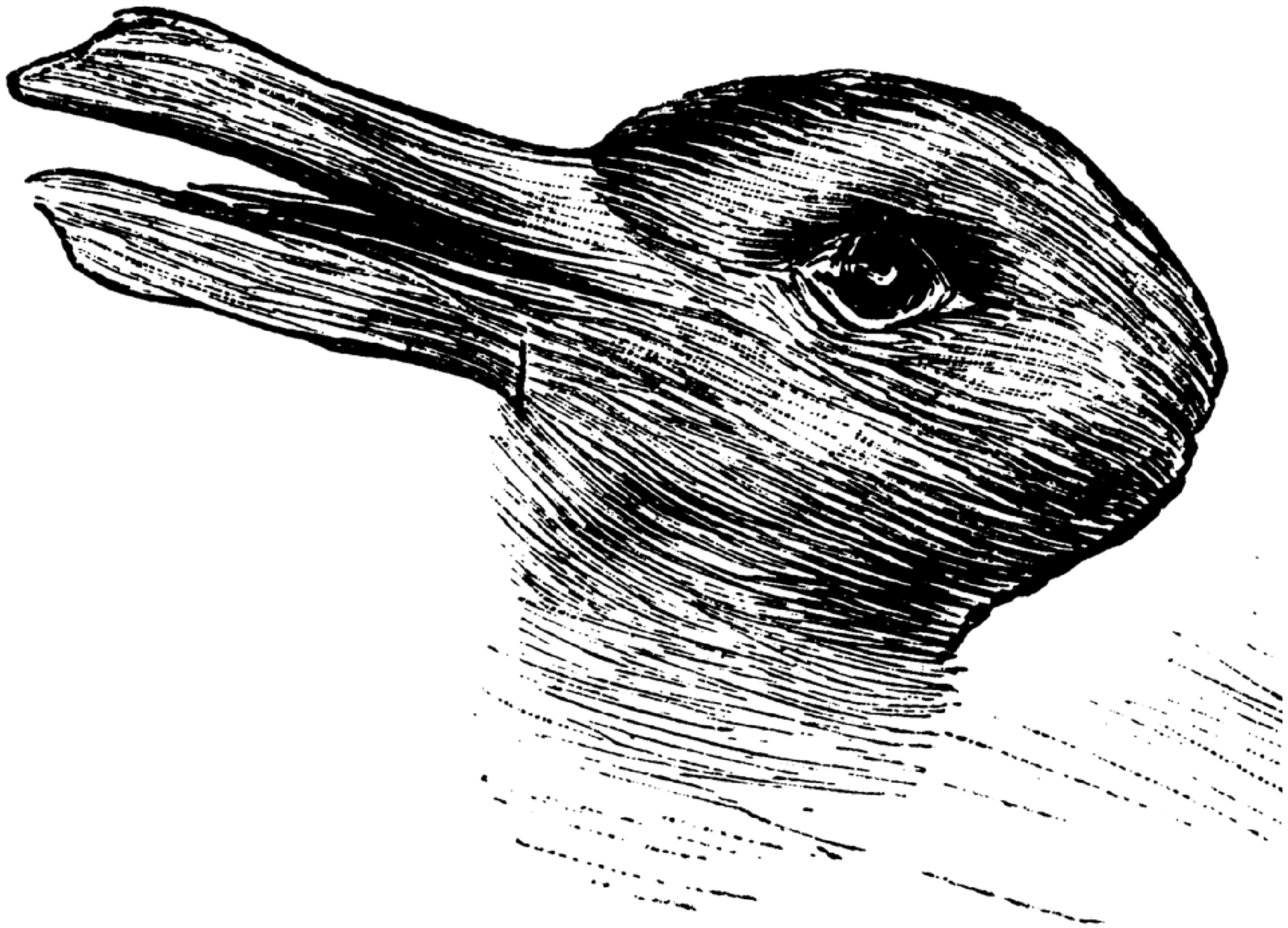
Lehky & Sejnowski (1988). Network model of shape-from-shading: neural function arises from both receptive and projective fields. *Nature*, 333: 452–454.

Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C., and Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature*, 435(7045):1102–1107.

Sherman, S.M. & Guillery, R.W. (2002). The role of the thalamus in the flow of information to the cortex. *Philosophical transactions of the Royal Society of London Series B*, 357(1428): 1695-708.

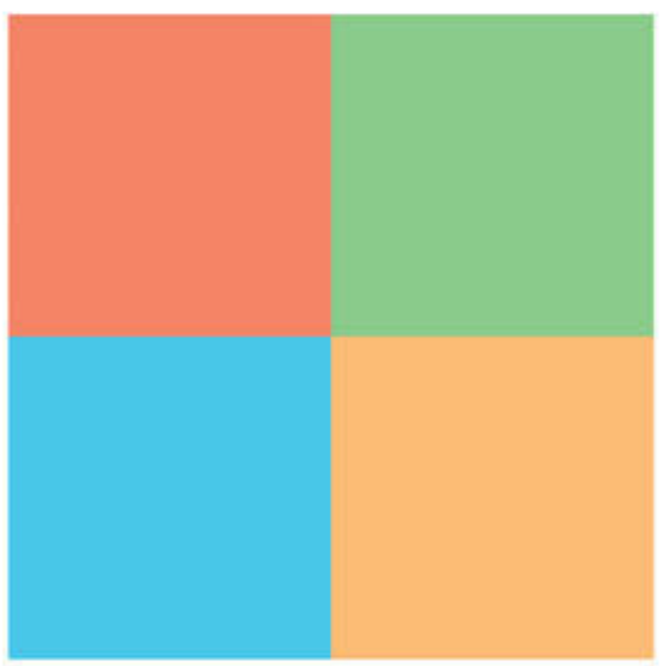
Walther, D. B. and Shen, D. (2014). Nonaccidental Properties Underlie Human Categorization of Complex Natural Scenes. *Psychological Science*, 25(4):851–860.

Mid-level vision (continued)



"There can be no peace until they renounce their Rabbit God and accept our Duck God."

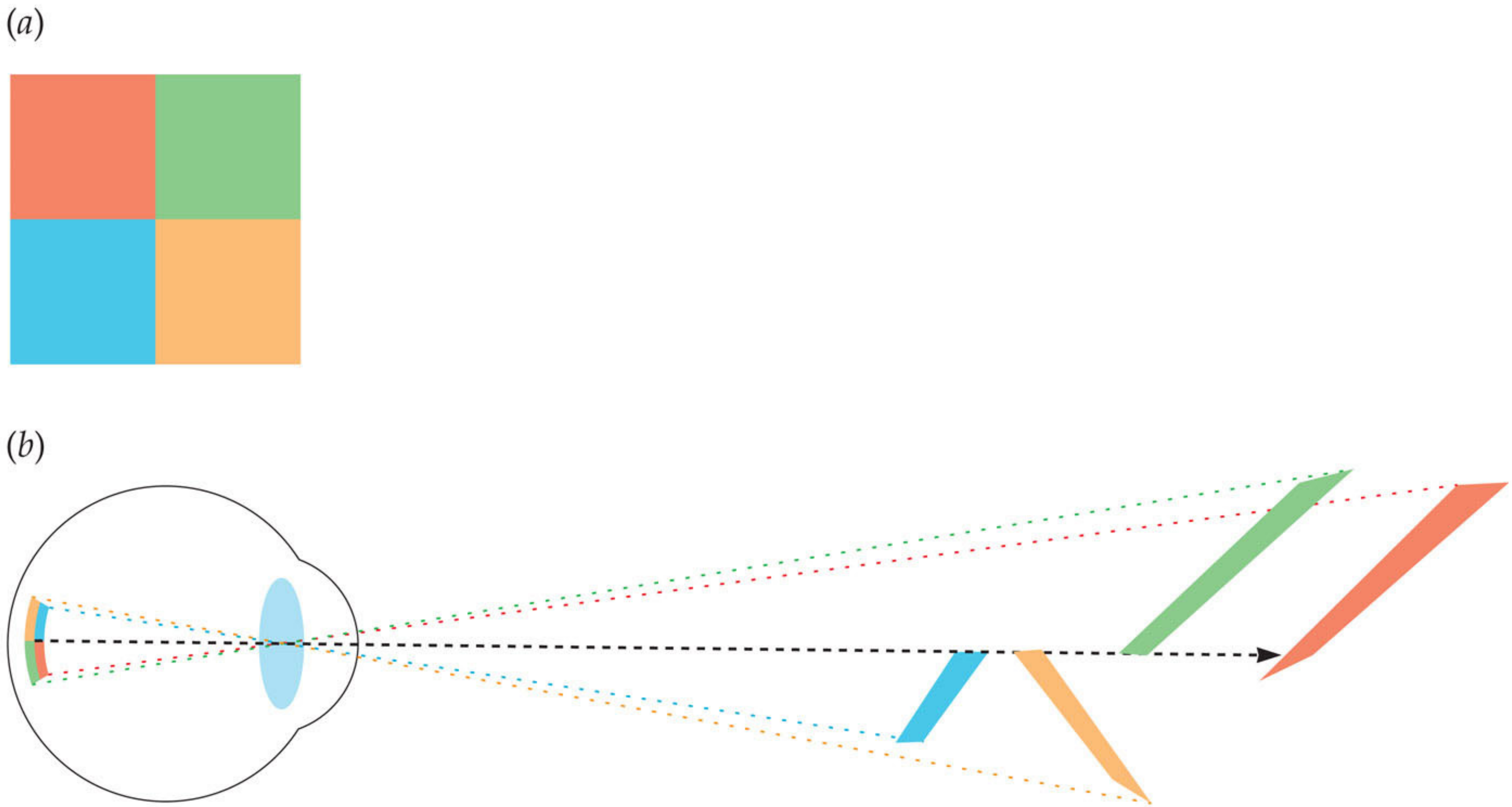
(*a*)



Mid-level vision: Accidental viewpoints

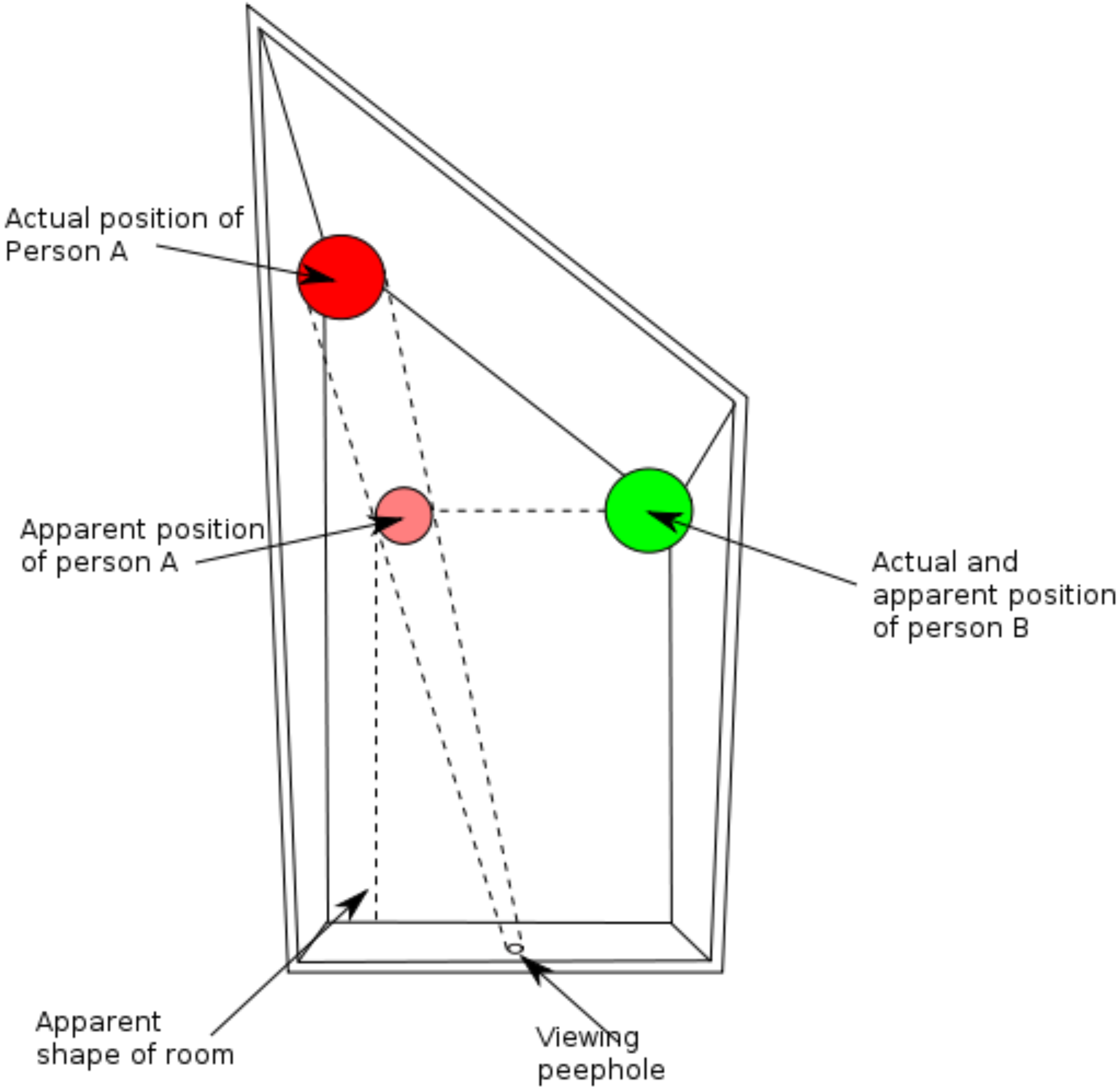
Accidental viewpoint: A viewing position that produces some regularity in the visual image that is not present in the world.

The visual system assumes viewpoints are not accidental.



***SENSATION & PERCEPTION 4e*, Figure 4.25**

https://c1.staticflickr.com/5/4049/5132242358_42134a1434_b.jpg



https://en.wikipedia.org/wiki/Ames_room

Forced perspective and the Beuchet chair

Forced perspective and the Beuchet chair

Forced perspective and the Beuchet chair

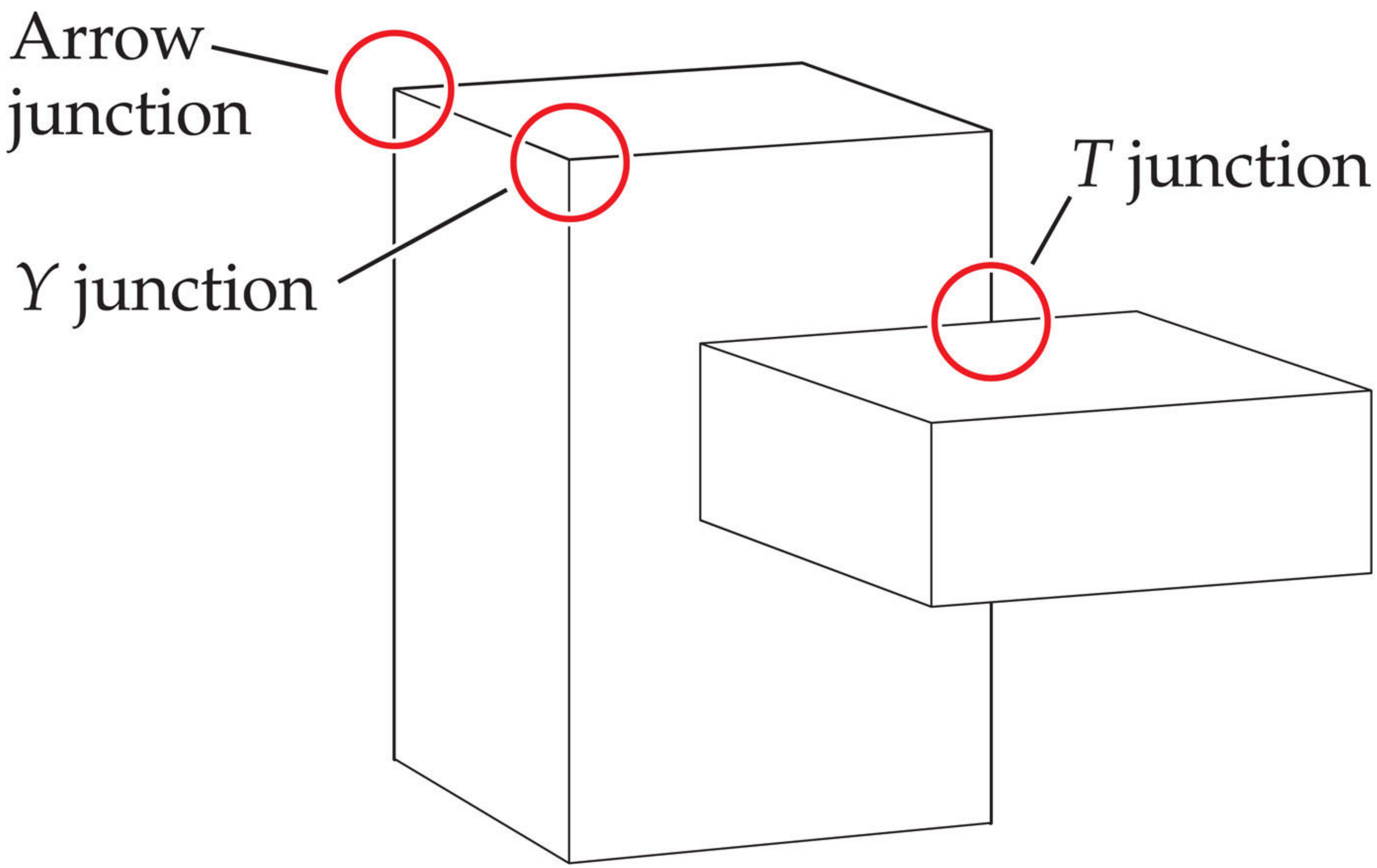
Mid-level vision: Non-accidental features

Non-accidental feature: A feature of an object that is not dependent on the exact (or accidental) viewing position of the observer (Witkin & Tenenbaum, 1983)

T junctions: Indicate occlusion. Top of T is in front and stem of T is in back.

Y junctions: Indicate corners facing the observer.

Arrow junctions: Indicate corners facing away from the observer.



SENSATION & PERCEPTION 4e, Figure 4.32

Mid-level vision: Non-accidental features

The central organizational principle is that certain properties of edges in a two-dimensional image are taken by the visual system as strong evidence that the edges in the three-dimensional world contain those same properties. For example, if there is a straight line in the image *(collinearity),* the visual system infers that the edge producing that line in the three-dimensional world is also straight. The visual system ignores the possibility that the property in the image might be a result of a (highly unlikely) accidental alignment of eye and curved edge.

Biederman, 1987

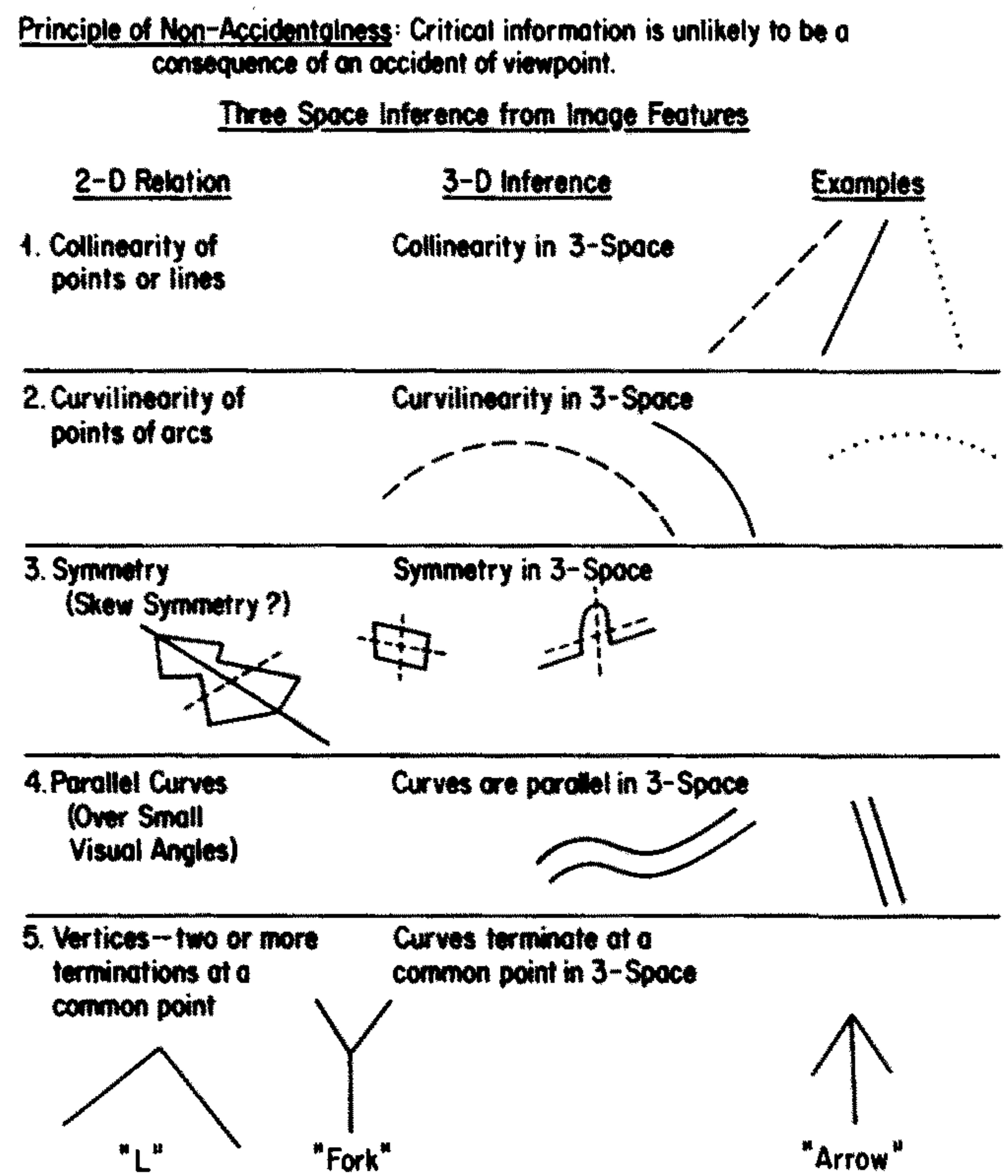
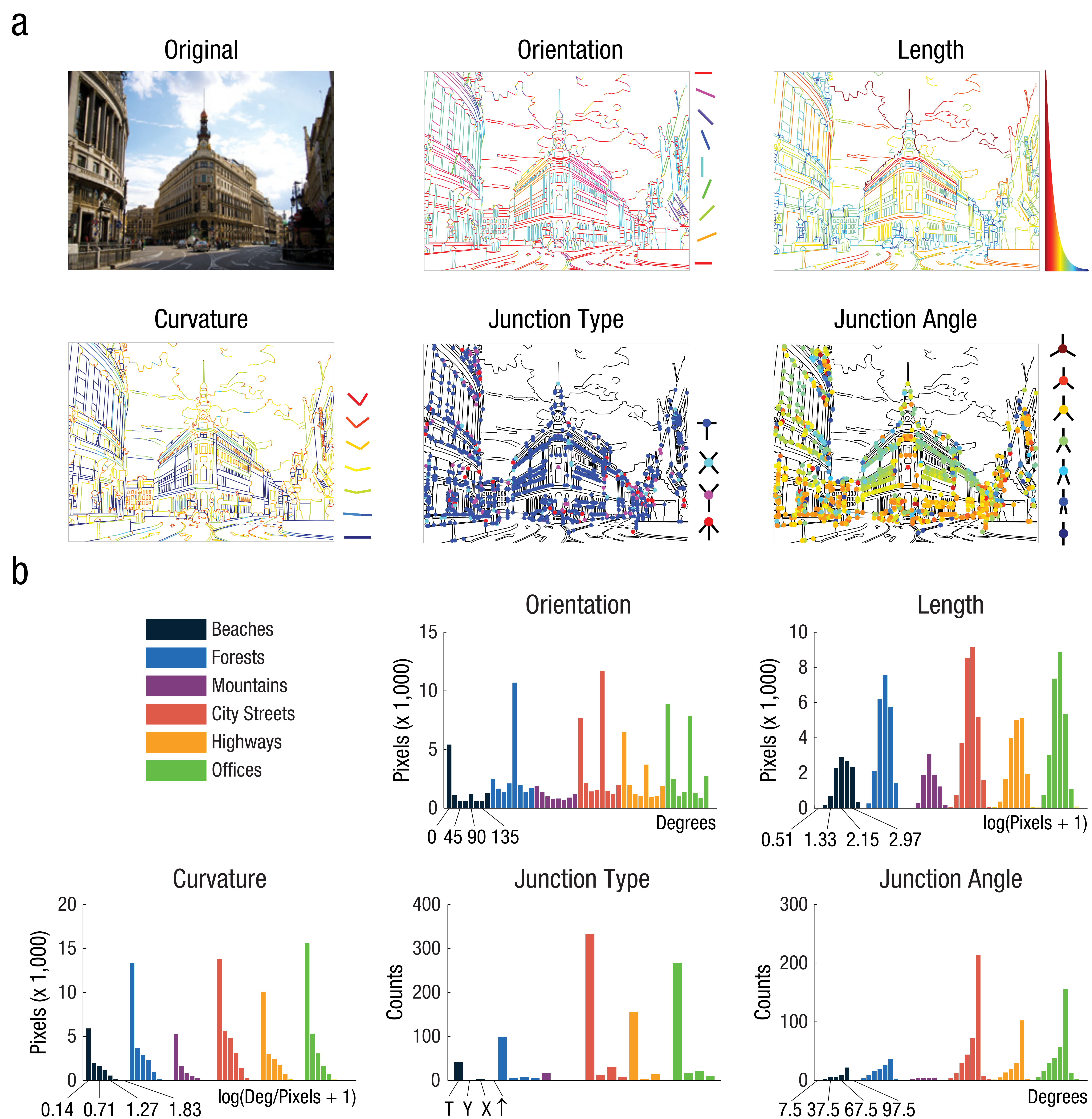


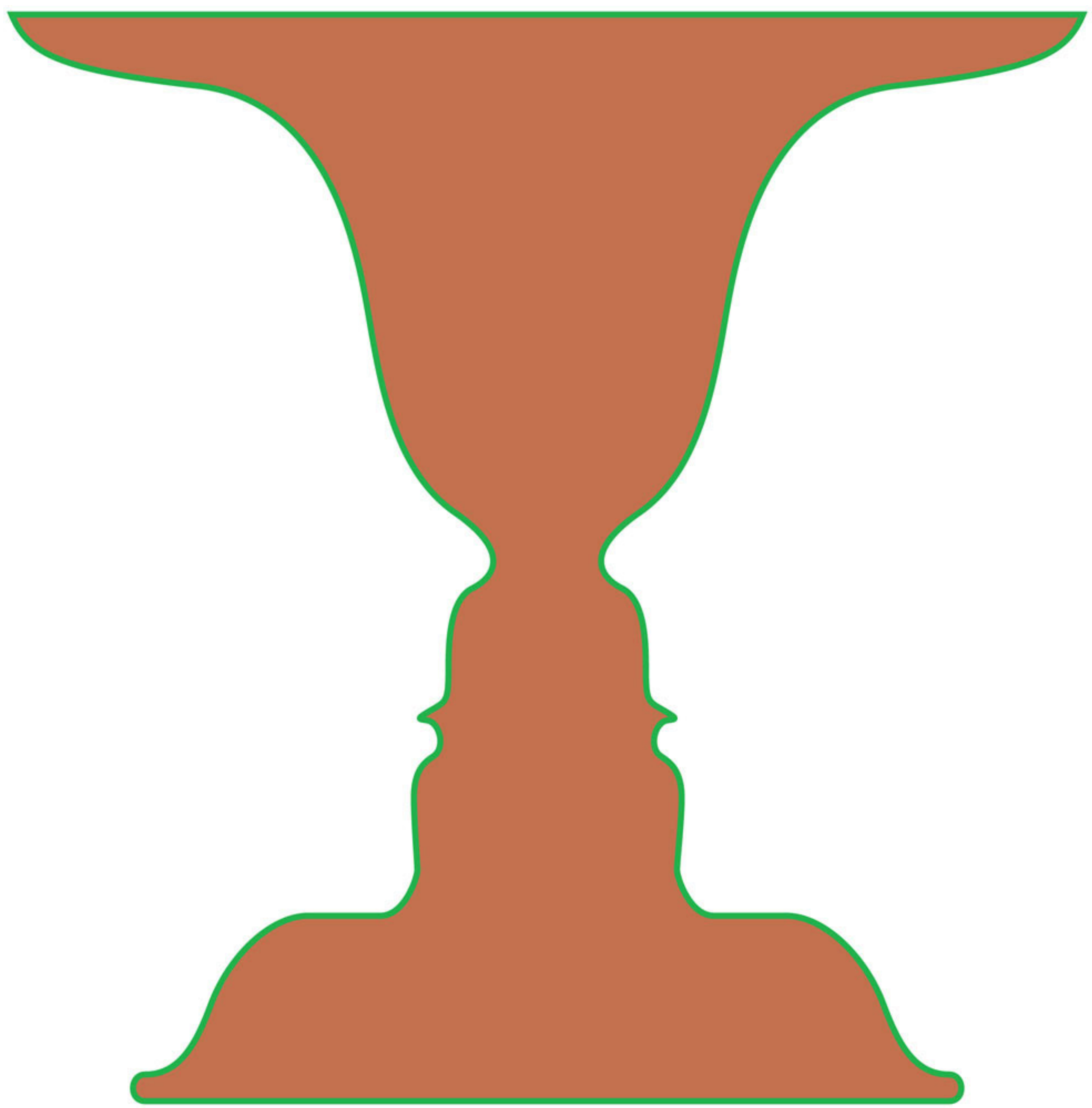
Figure 4. Five nonaccidental relations. (From Figure 5.2, *Perceptual organization and visual recognition* [p. 77] by David Lowe. Unpublished doctorial dissertation, Stanford University. Adapted by permission.)



From: Walther, D. B., & Shen, D. (2014)

Mid-level vision: Figure-ground

Figure-ground assignment: The process of determining that some regions of an image belong to a foreground object (figure) and other regions are part of the background (ground).



SENSATION & PERCEPTION 4e, **Figure 4.27**

Mid-level vision: Gestalt figure-ground assignment

Surroundedness: The surrounding region is likely to be ground.

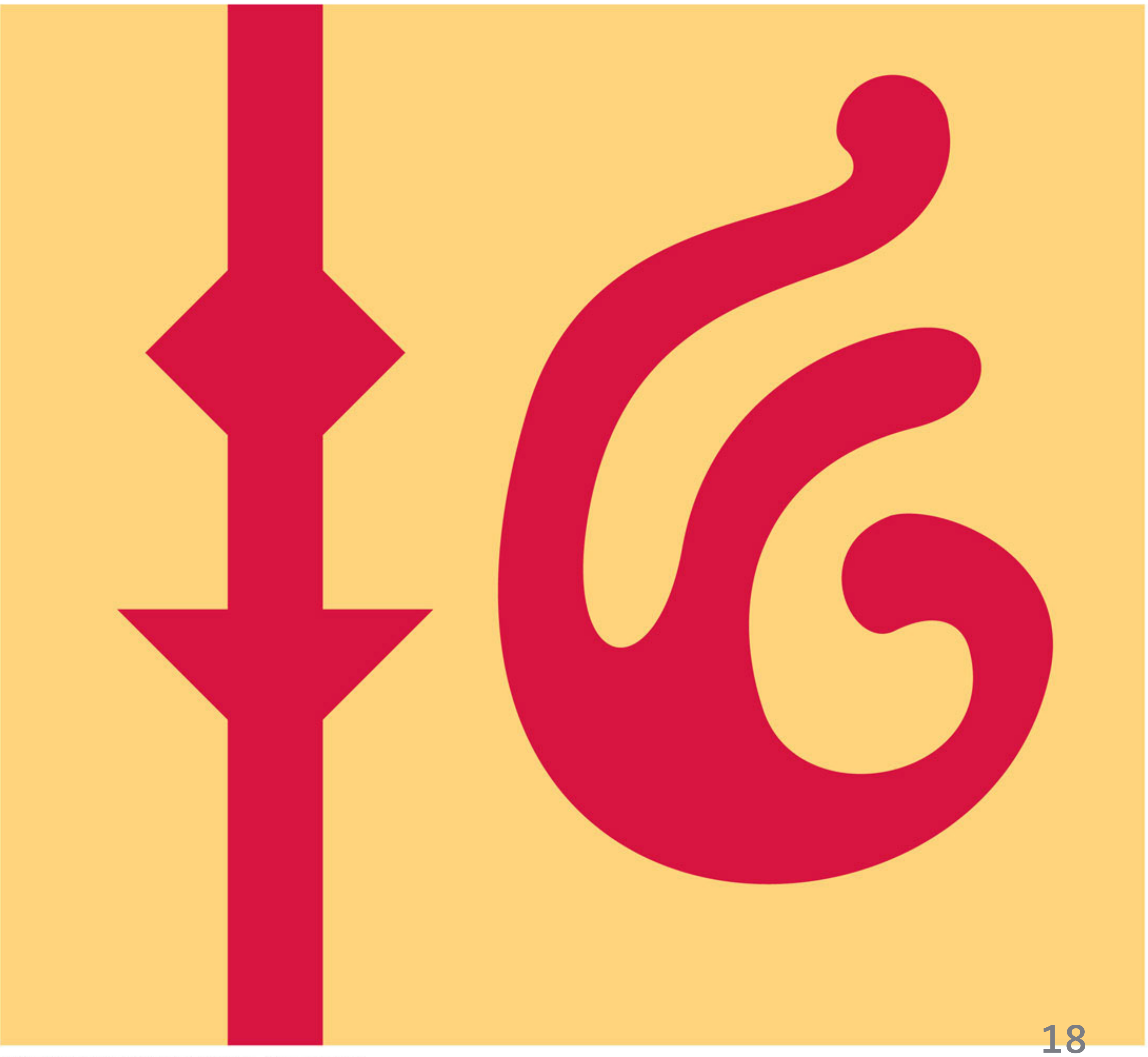
Size: The smaller region is likely to be figure.

Symmetry: A symmetrical region tends to be seen as figure.

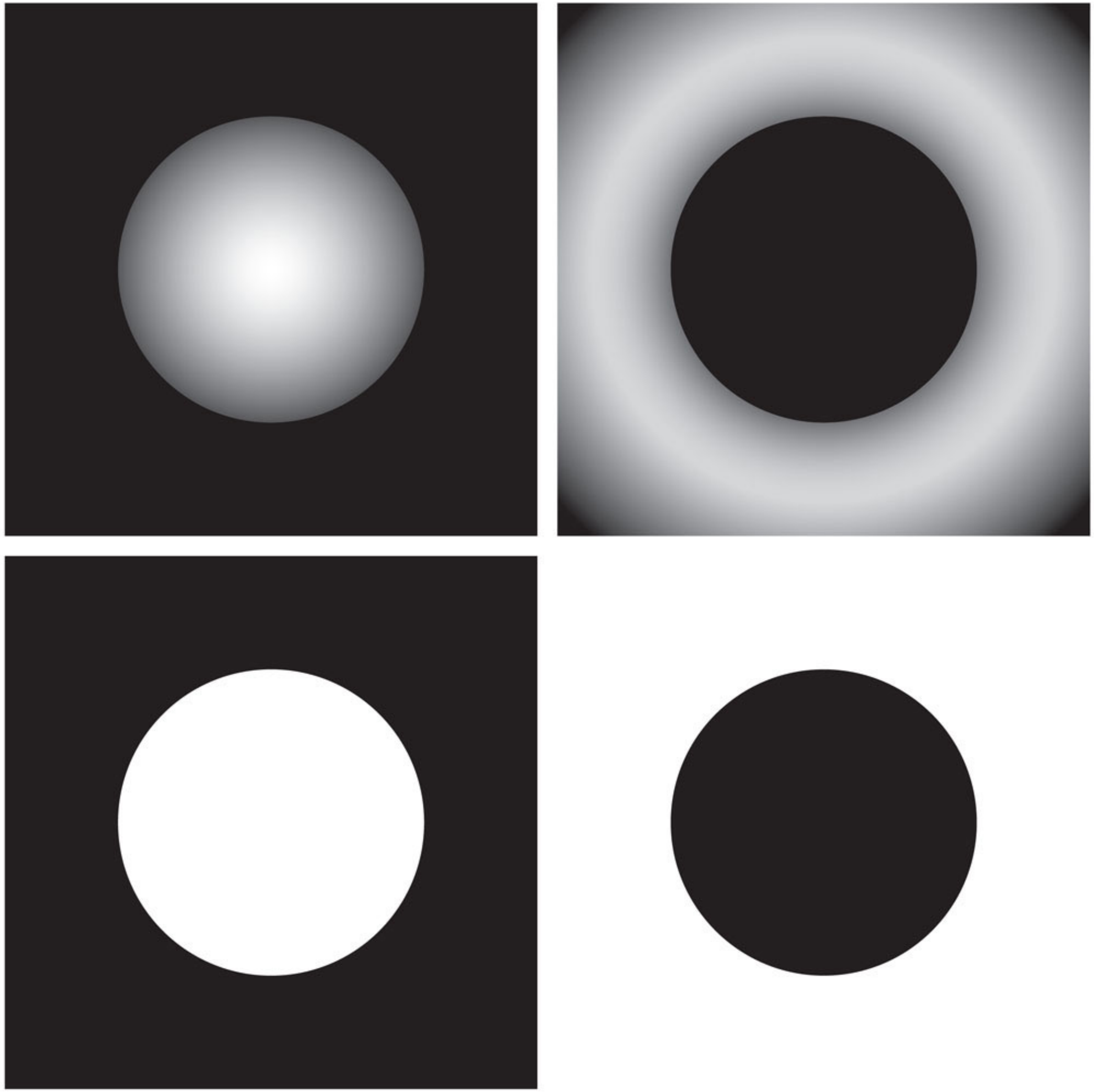
Parallelism: Regions with parallel contours tend to be seen as figure.

Extremal edges: If edges of an object are shaded such that they seem to recede in the distance, they tend to be seen as figure.

Relative motion: If one region moves in front of another, then the closer region is figure.

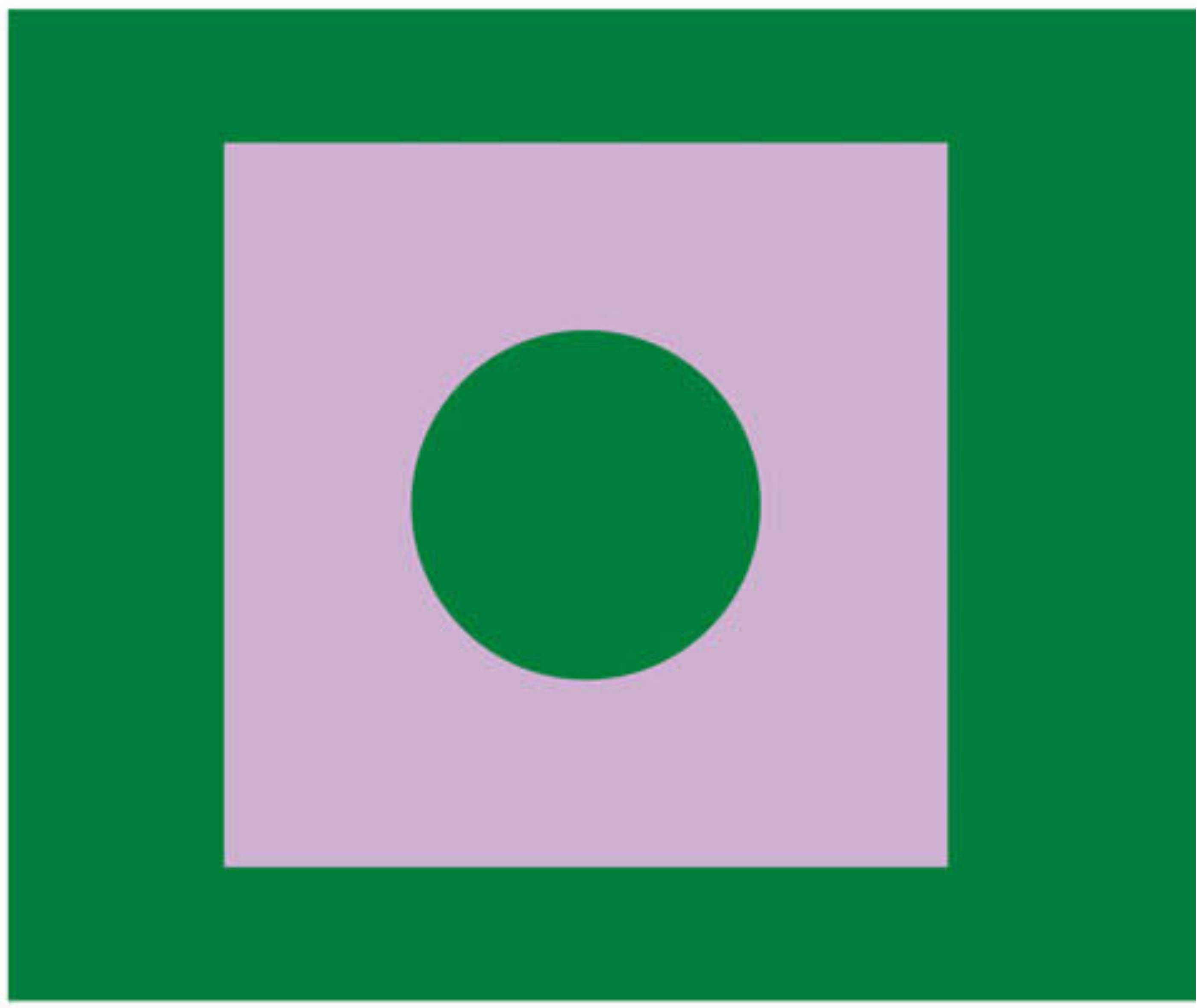


SENSATION & PERCEPTION 4e, Figure 4.26



***SENSATION & PERCEPTION 4e,* Figure 4.29**

(a)

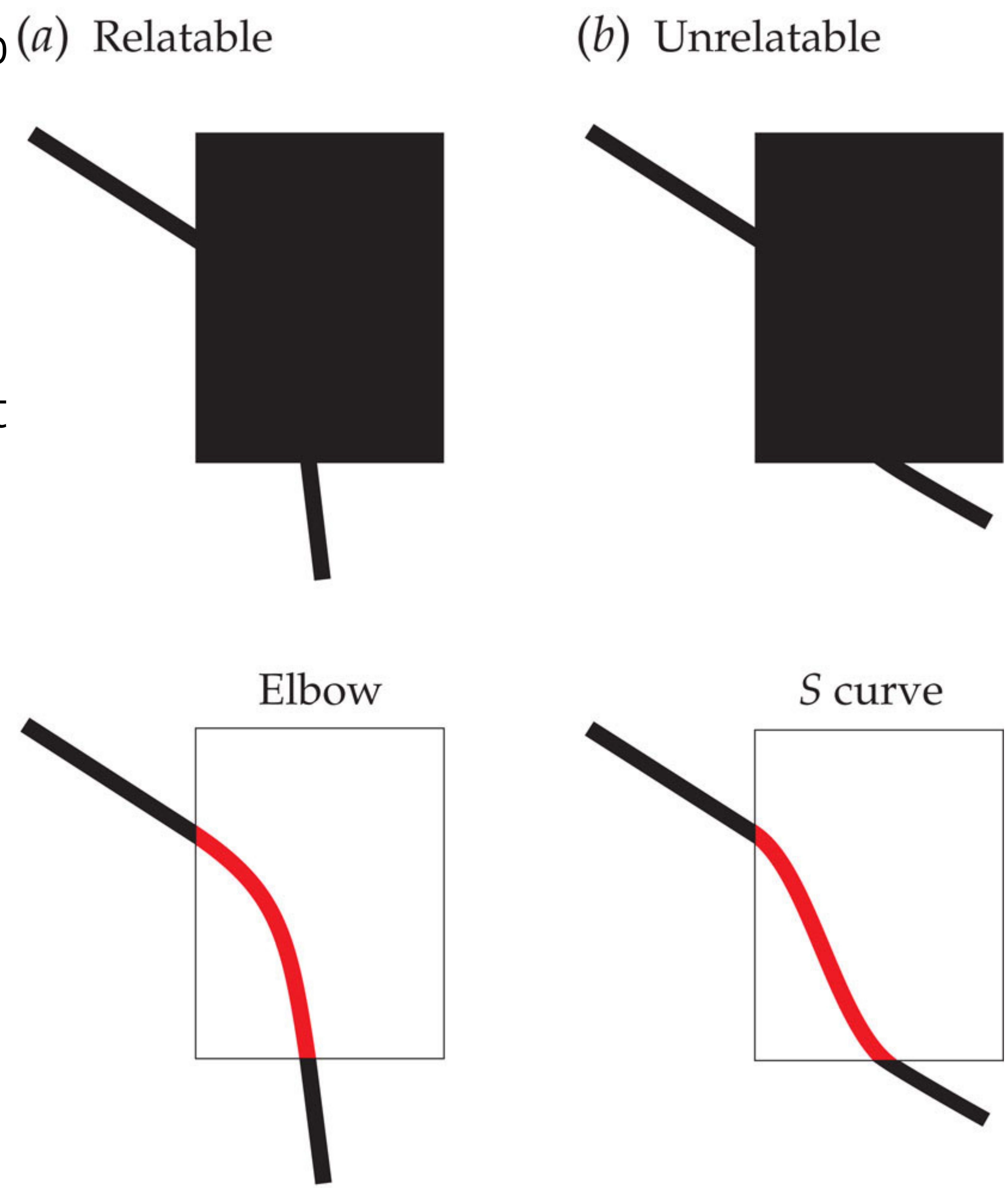


***SENSATION & PERCEPTION 4e,* Figure 4.30**

Mid-level vision: Occlusion

Relatability: The degree to which two line segments appear to be part of the same contour.

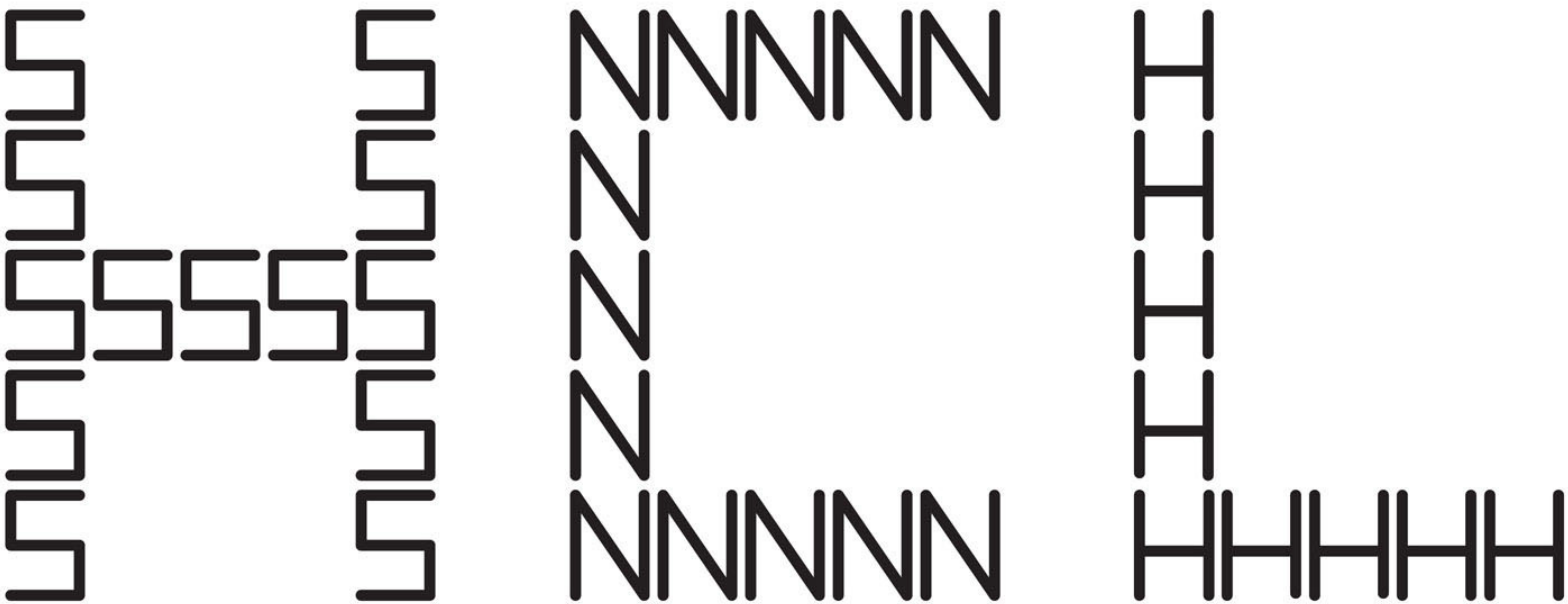
Two edges are relatable if they can be connected with a smooth convex or smooth concave curve (a), but not if the connection requires an S curve (b)



SENSATION & PERCEPTION 4e, **Figure 4.31**

Mid-level vision: Parts and wholes

Global superiority effect: The properties of the whole object take precedence over the properties of parts of the object.



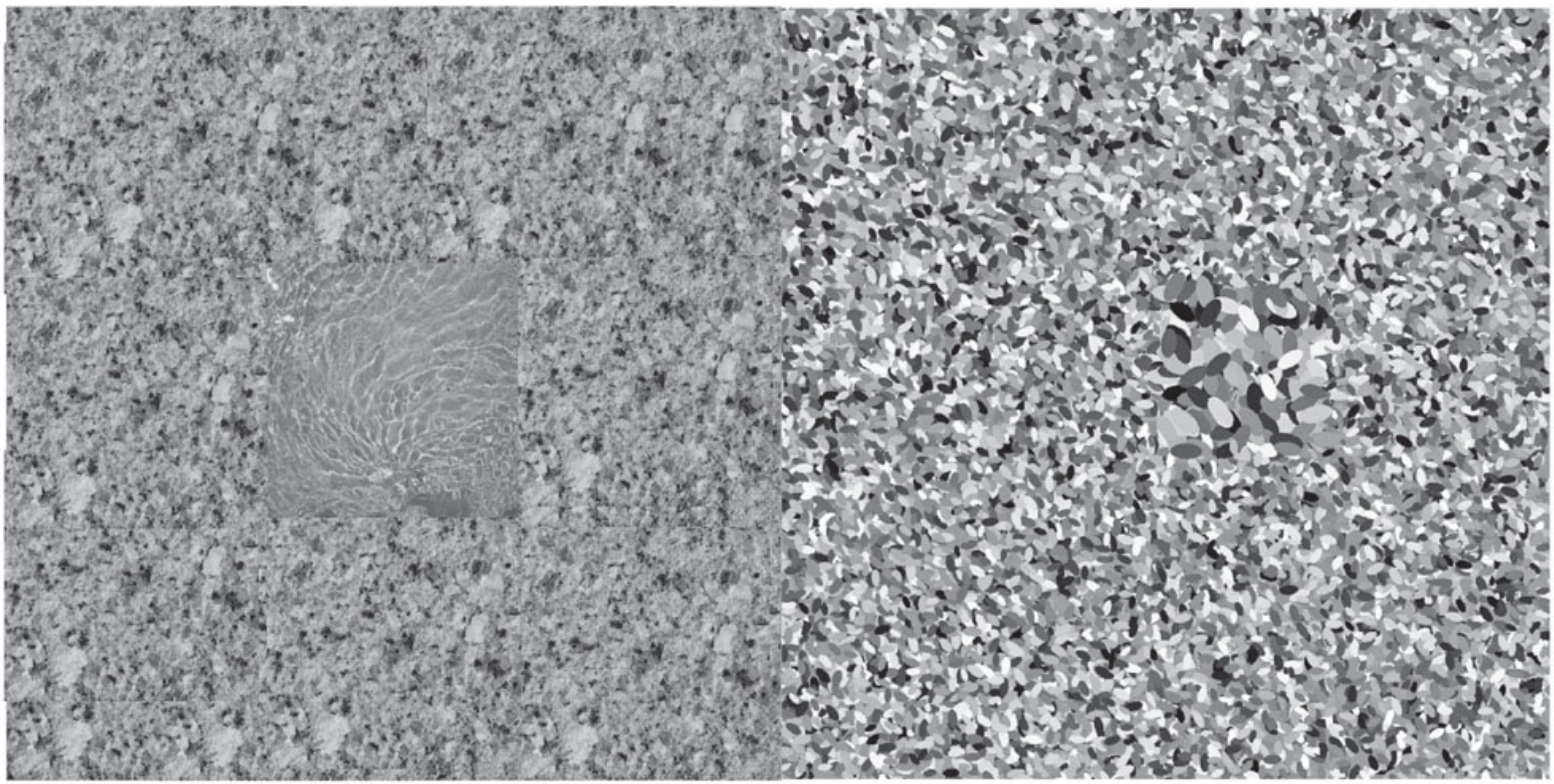
SENSATION & PERCEPTION 4e, **Figure 4.33**

Navon, 1977

Mid-level vision: Texture segmentation and grouping

Texture segmentation: Carving an image into regions of common texture properties.

Texture grouping depends on the statistics of textures in one region versus

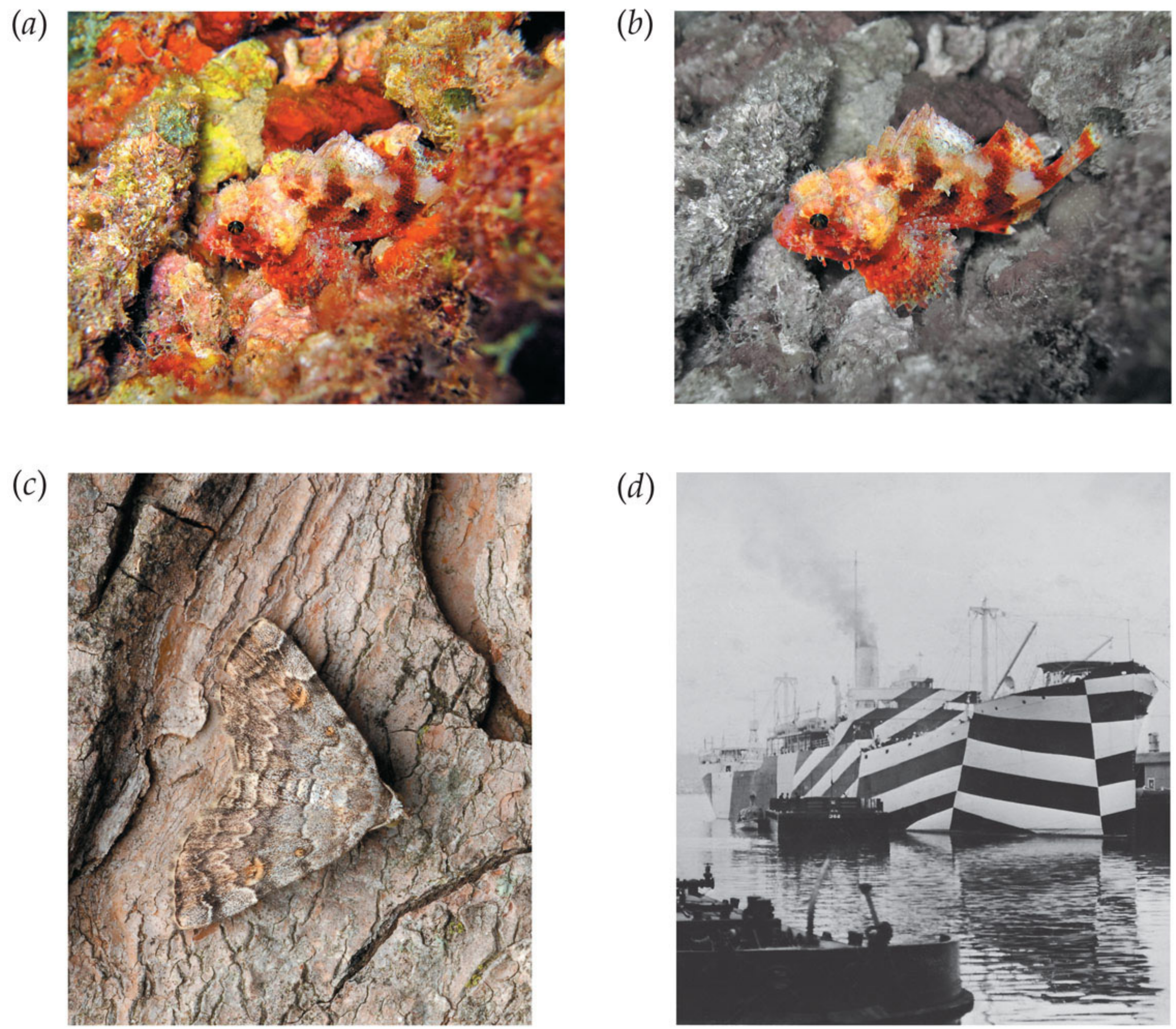


SENSATION & PERCEPTION 4e, Figure 4.16

Mid-level vision: Camouflage

Animals exploit Gestalt grouping principles to group into their surroundings.

Sometimes camouflage is used to confuse the observer.



SENSATION & PERCEPTION 4e, **Figure 4.21**

Mid-level vision: Summary

Five principles of middle vision:

1. Bring together that which should be brought together
2. Split asunder that which should be split asunder
3. Use what you know
4. Avoid accidents
5. Seek consensus and avoid ambiguity

Bayesian approaches to perception

The "committee" metaphor has been formalised by those who propose that the visual system performs (approximate) Bayesian inference:

Prior: what you believe beforehand. Knowledge of mid-level principles comes here.

Likelihood: how consistent is the input with all possible hypotheses?

Posterior: combination of likelihood and prior.

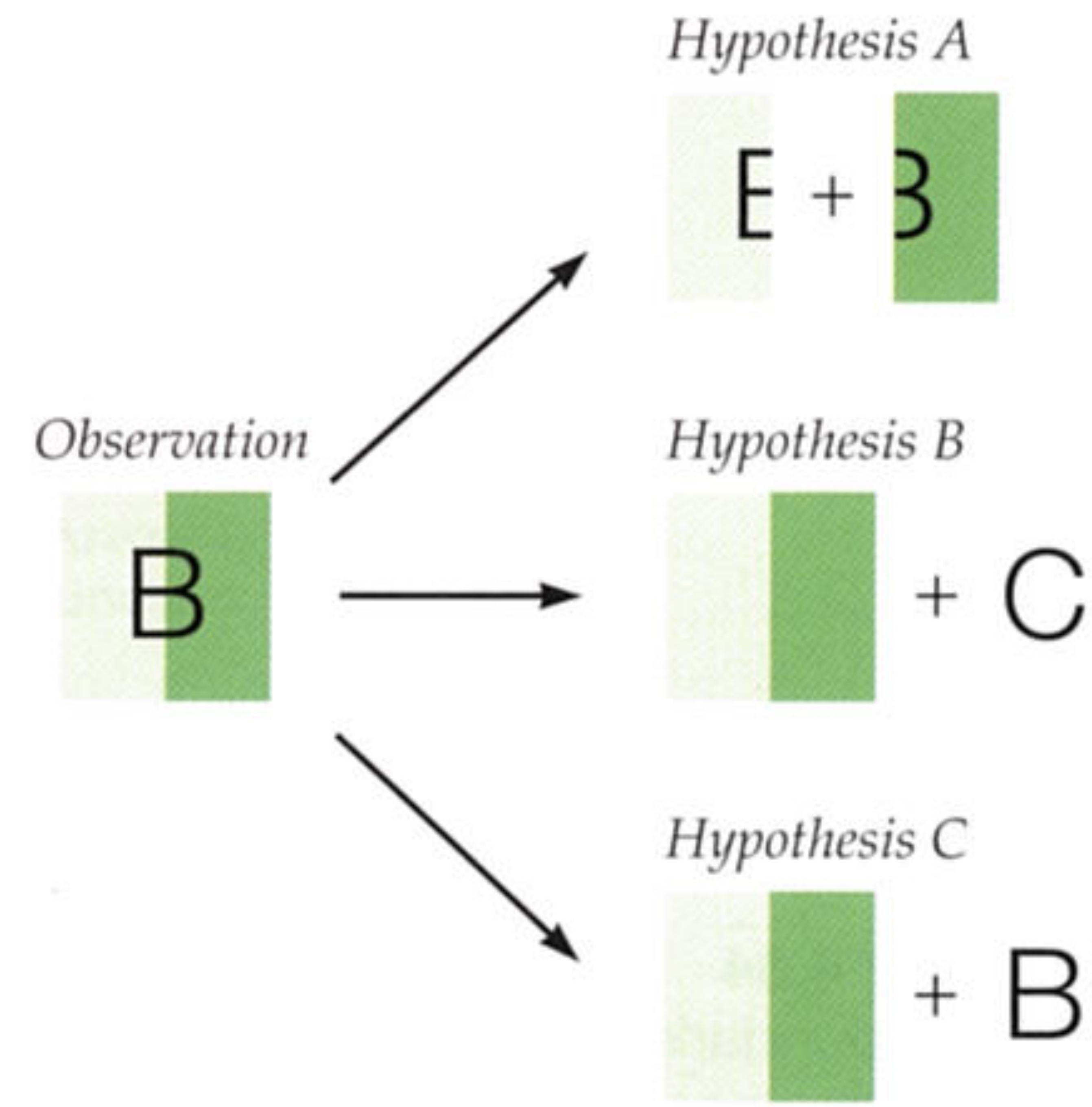
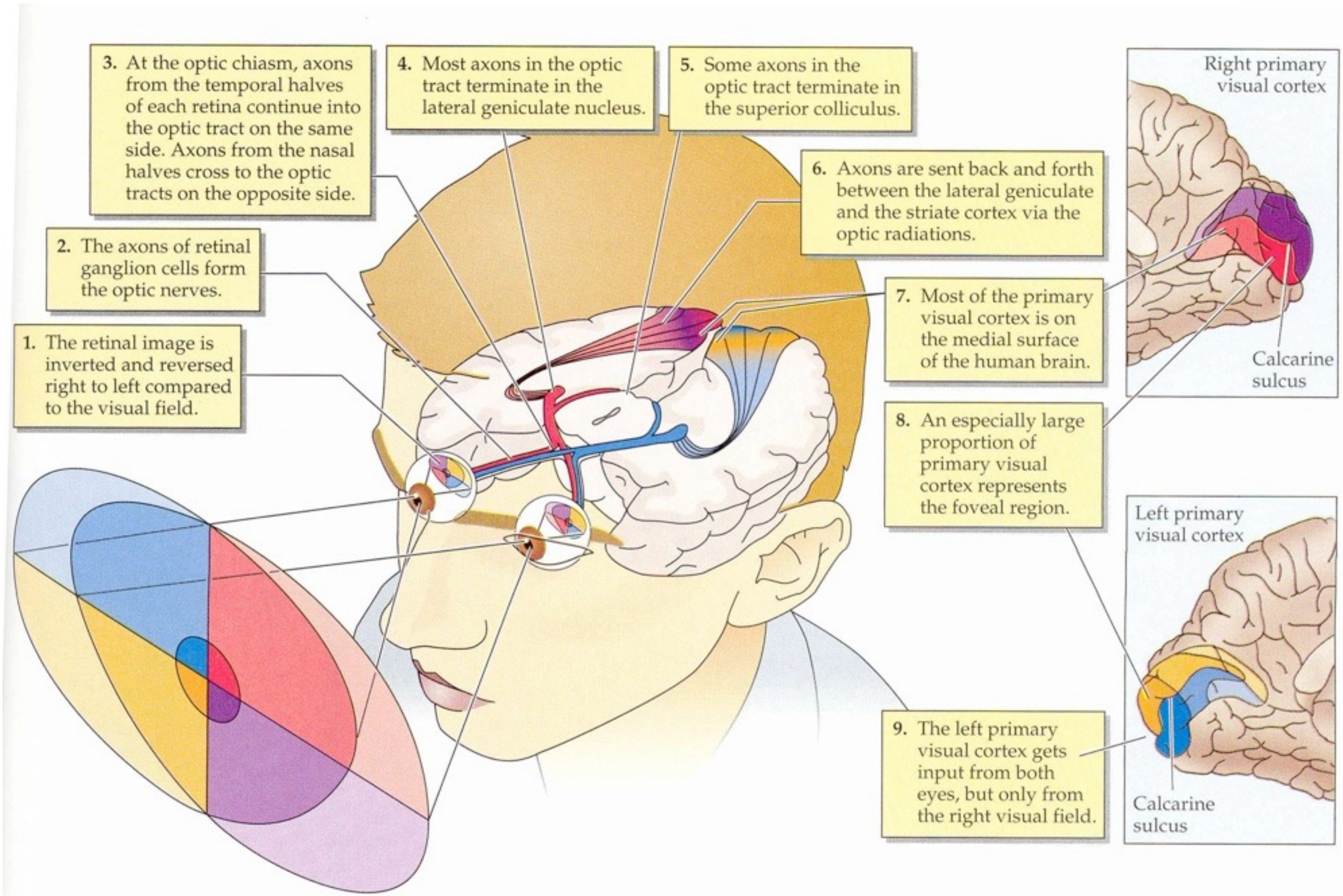


FIGURE 4.35 When the visual system is faced with a stimulus, it tries to figure out the most likely situation in the world that has produced this particular pattern of activity.

Neuroscience of object recognition

Retino-Cortical Pathway



From spots and bars to objects and space

How does the brain recognise objects?

Retinal ganglion cells = "spots"

V1 = "bars" (local, blurred, derivative operators – "stuff")

Clearly our brains are doing something pretty sophisticated beyond V1.

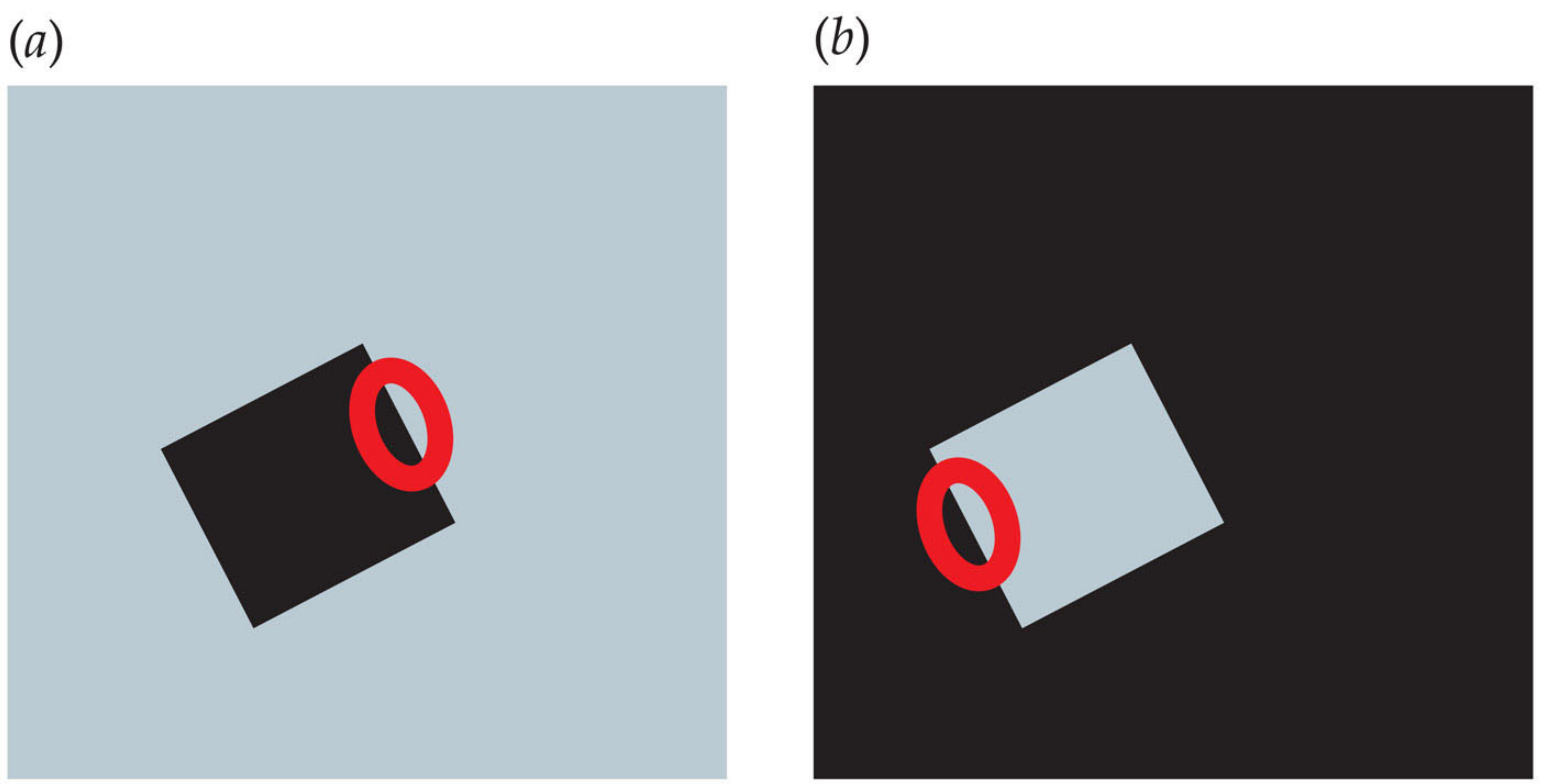
Receptive fields in extrastriate areas

The receptive fields of extrastriate cells are more sophisticated than those in striate cortex, responding to more complex properties.

For instance, "border ownership." For a given border, which side is part of the object and which side is part of the background?

The same visual input occurs in both (a) and (b) and a V1 neuron would respond equally to both.

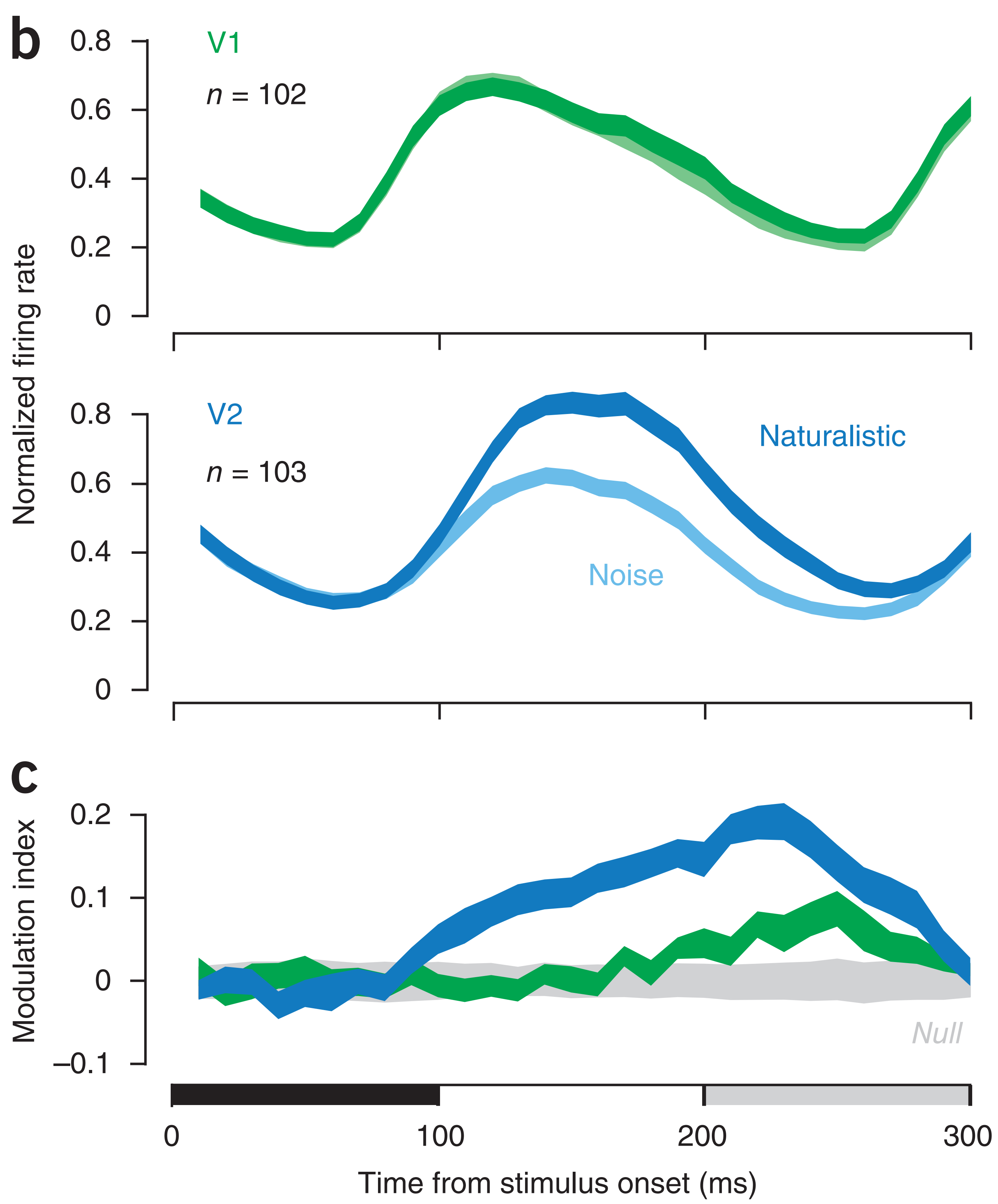
A V2 neuron might respond more to (a) than (b) because the black edge is owned by the square in (a) but not in (b).

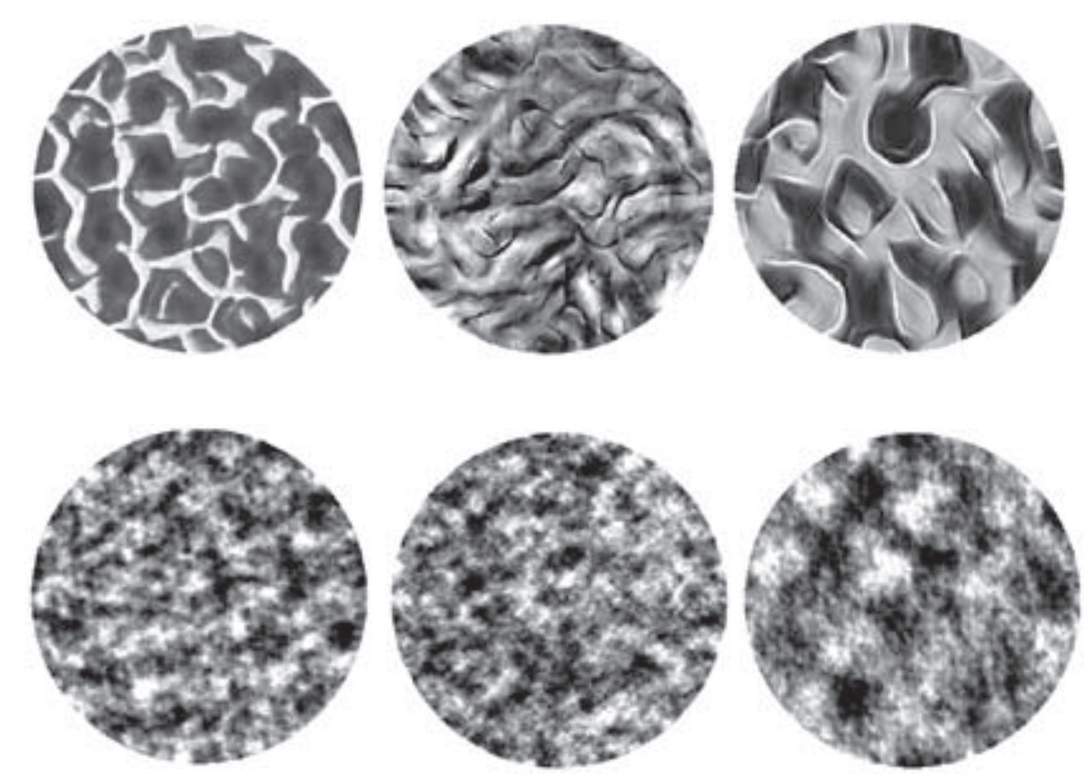


SENSATION & PERCEPTION 4e, **Figure 4.5**

Receptive fields in extrastriate areas

Another stimulus dimension that V2 seems to be selective for is texture: V2 neurons respond more to "naturalistic" texture stimuli than to amplitude-matched noise, whereas V1 neurons don't differentiate the two:





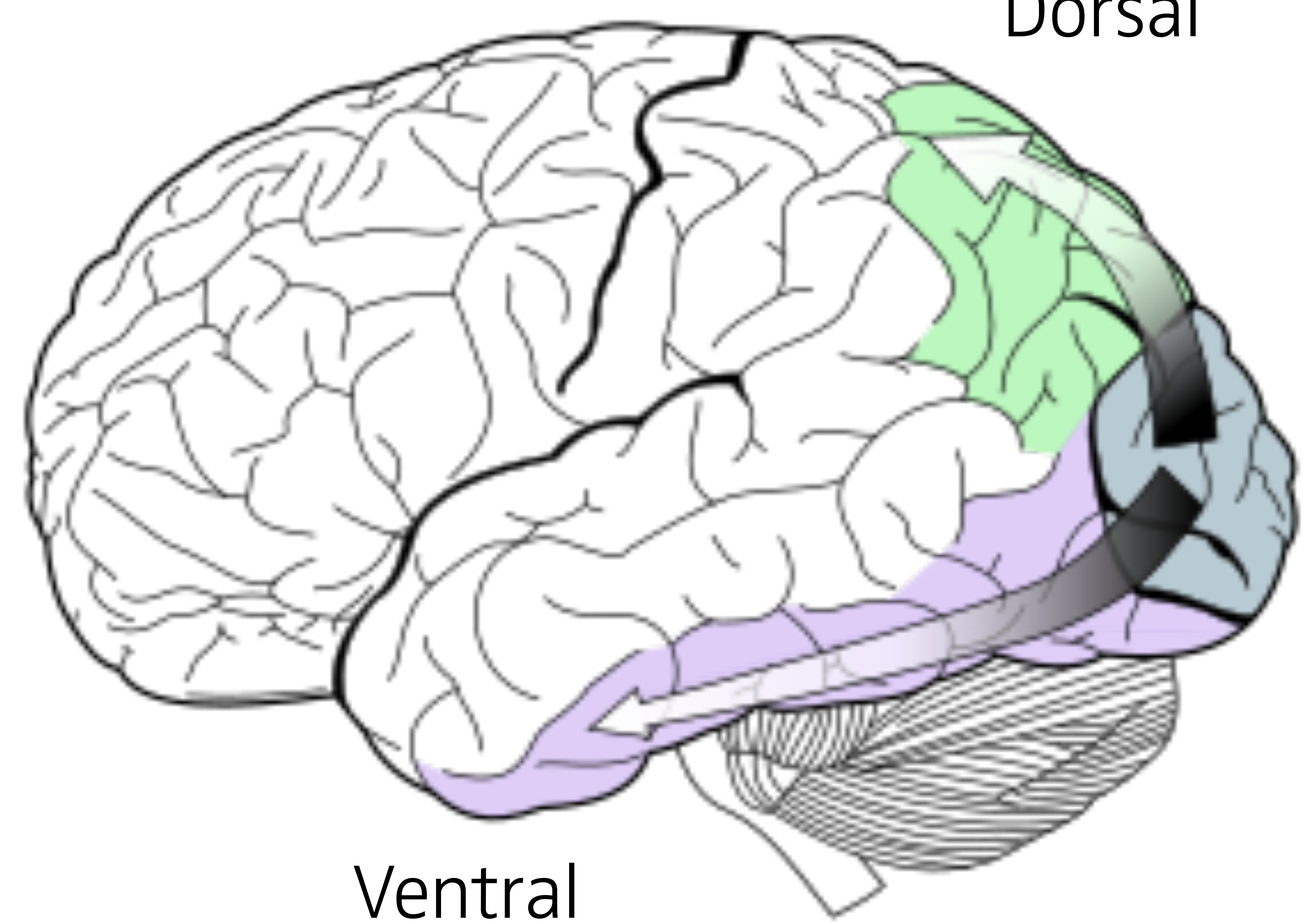
From Freeman, J., Ziemba, C., Heeger, D. J., Simoncelli, E. P., & Movshon, J. A. (2013). A functional and perceptual signature of the second visual area in primates. *Nature Neuroscience*, *16*(7), 974–981.

The two streams hypothesis (“what” and “where” pathways)

The visual cortex is thought to be organised into two relatively-specialised processing streams.

The *what* pathway (ventral stream) exhibits relative specialisation for object recognition: shapes, names and functions.

The *where* pathway (dorsal stream) shows relative specialisation for spatial localisation and the guidance of action.



https://en.wikipedia.org/wiki/Two-streams_hypothesis

From http://www.scholarpedia.org/article/What_and_where_pathways

Evidence for the two-streams hypothesis

Vast body of work from neuroanatomical, electrophysiological, and lesion studies supporting this view.

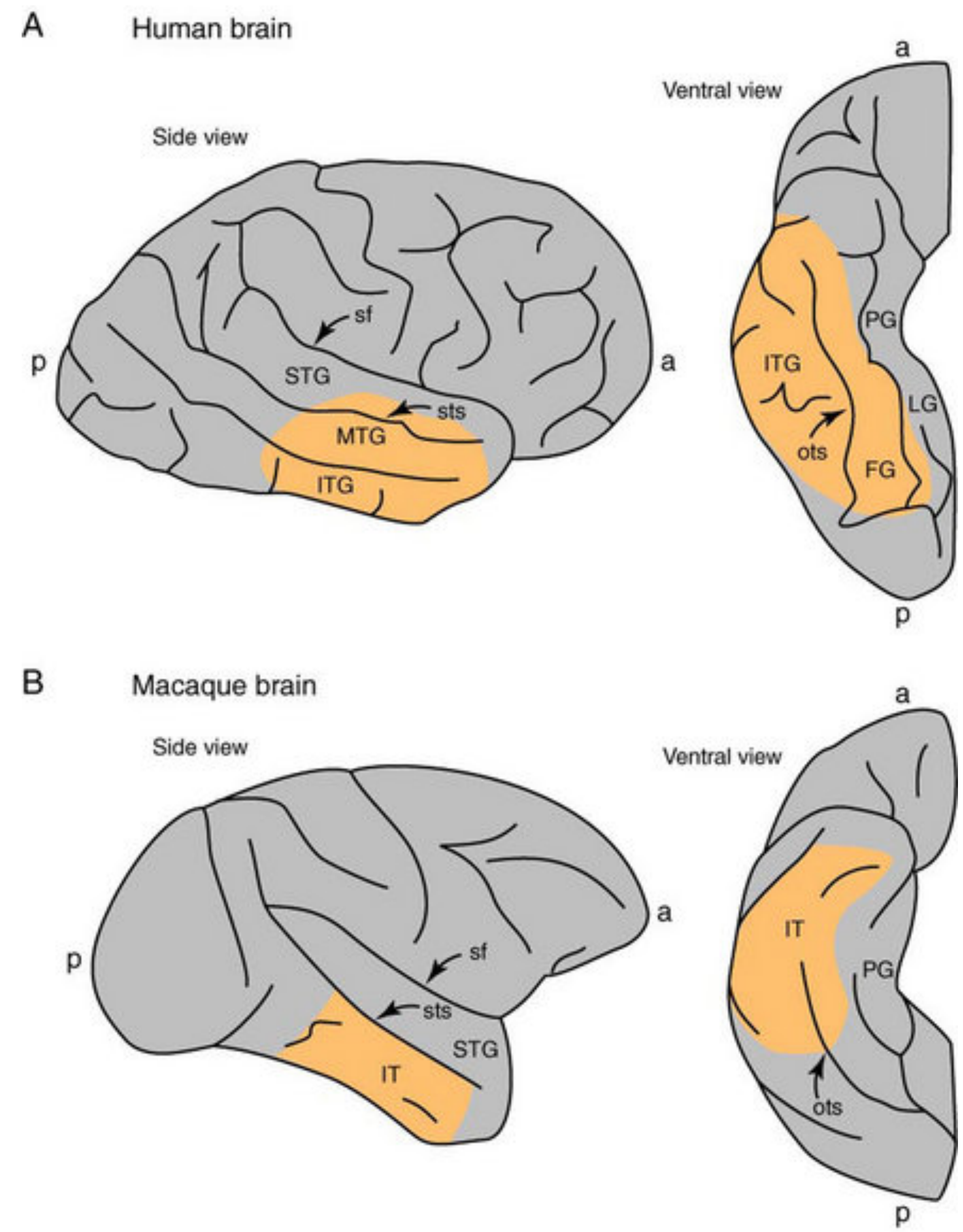
Earliest cues from studying humans with localised brain damage (lesions) due to e.g. stroke.

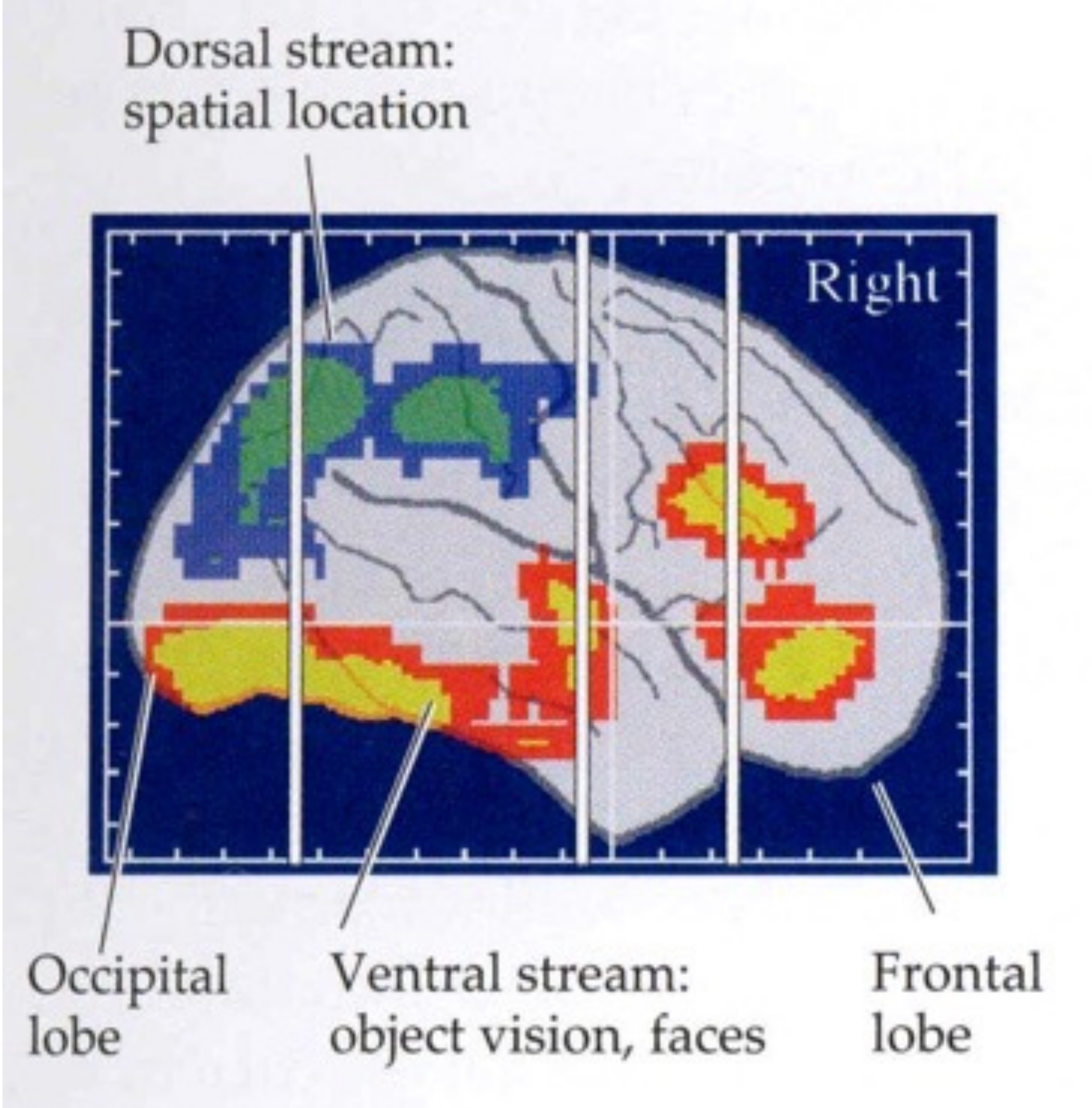
Lesion:

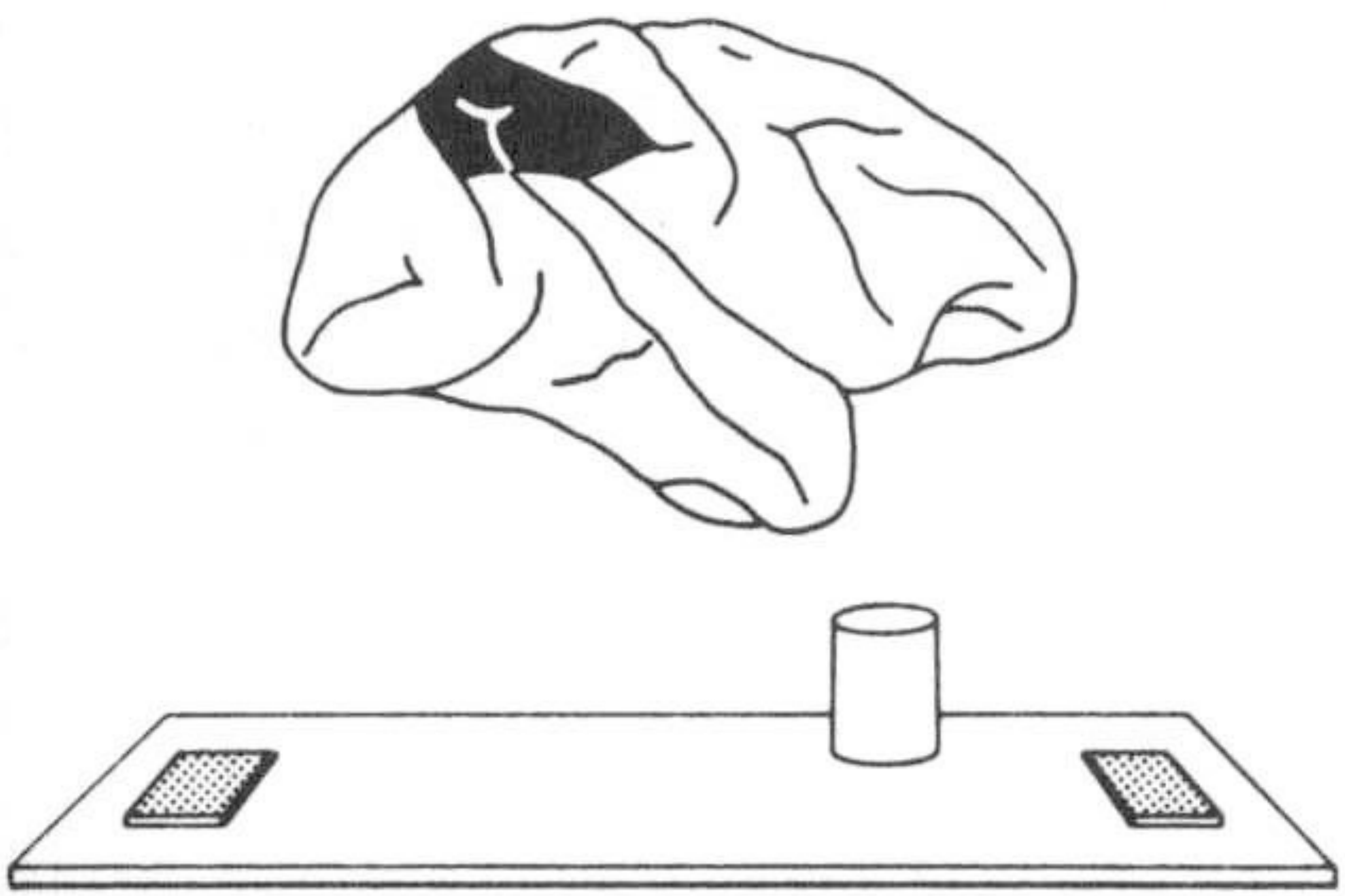
(n.) A region of damaged brain.
(v.) To destroy a section of the brain.

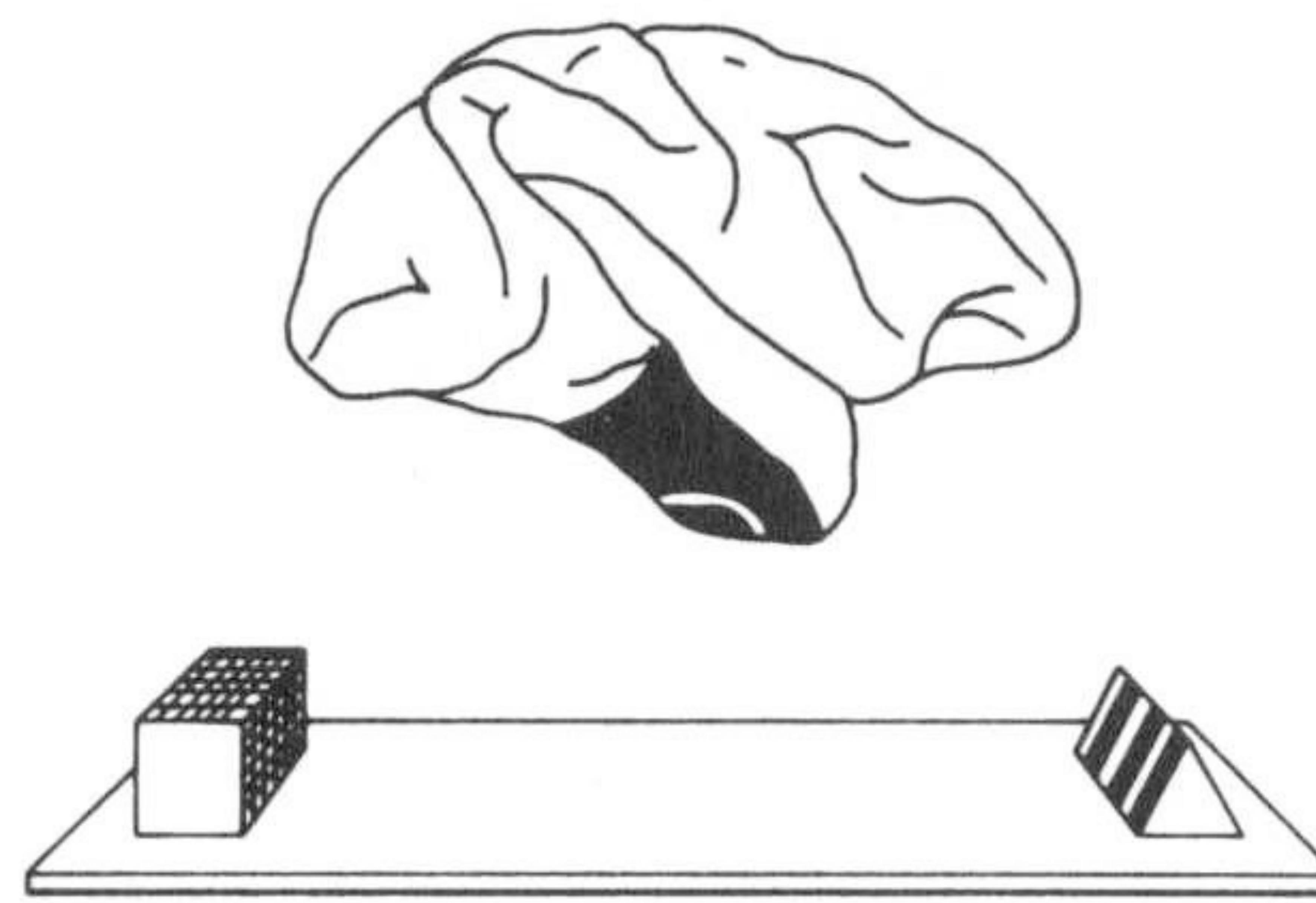
The inferior temporal (IT) cortex lies at the ''end'' of the ventral stream.

When IT cortex is lesioned, it leads to agnosias.



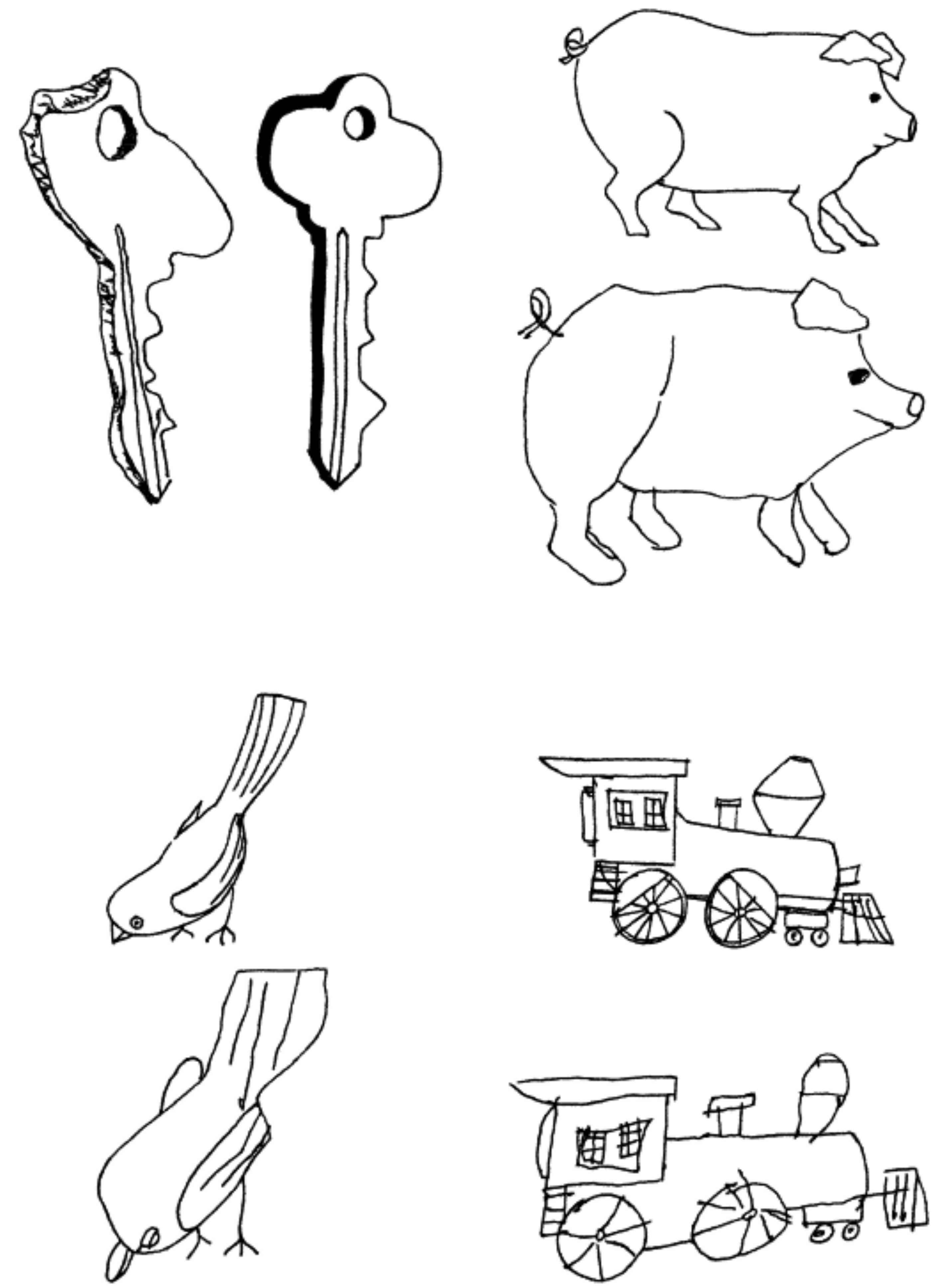

Dorsal stream:
spatial location
Right
Occipital
lobe
Ventral stream:
object vision, faces
Frontal
lobe





Agnosia

Failure to recognise objects in spite of the ability to see them (e.g. patients can copy image but not name the object).



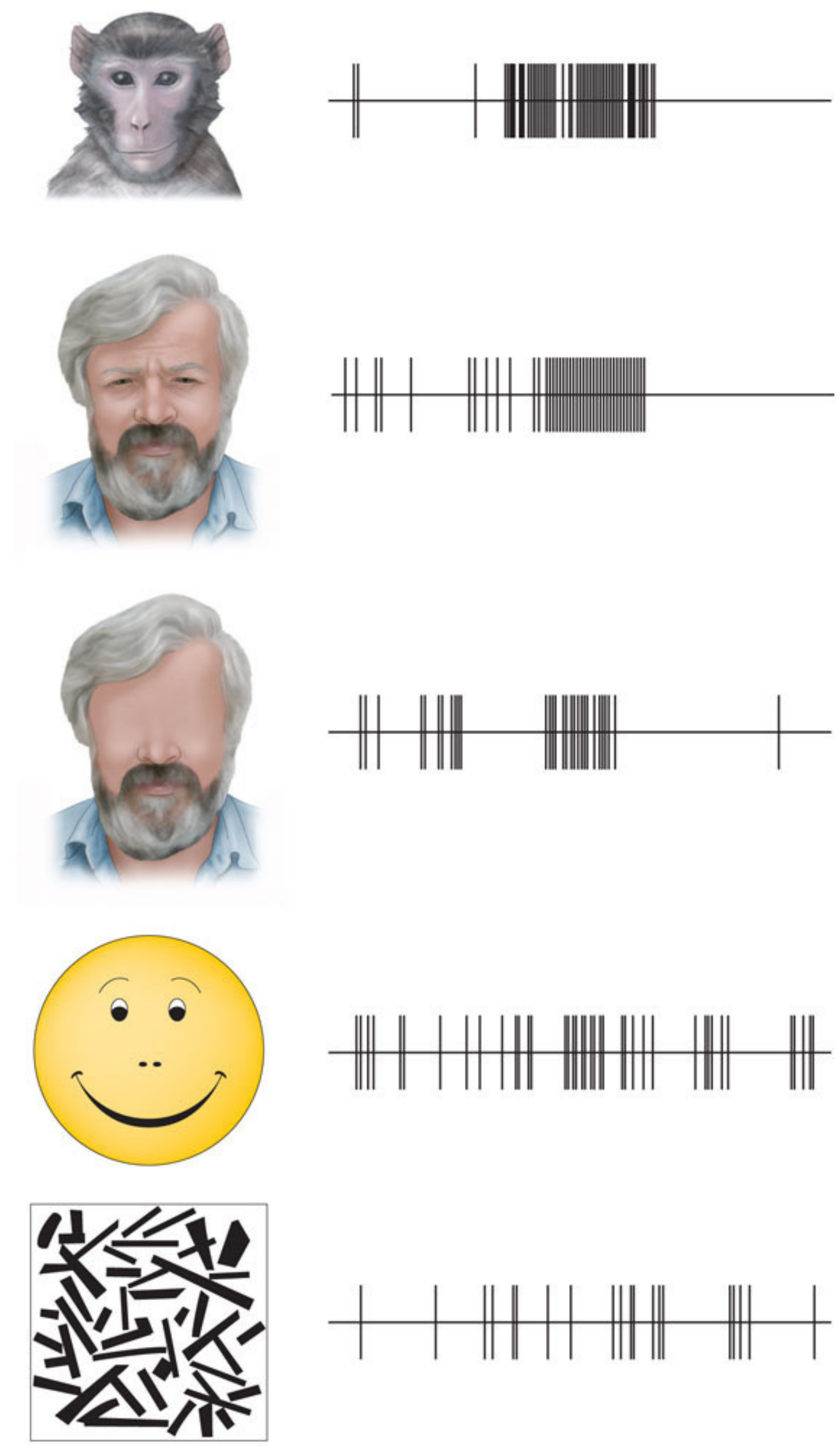
Lesioning IT cortex in monkeys leads to visual deficits with discrimination and recognition of objects analogous to agnosia in humans.

e.g. Prosopagnosia: inability to recognise faces

Receptive fields of IT neurons

Receptive field properties of IT neurons:

- Very large—some cover half the visual field
- Don't respond well to spots or lines
- Do respond well to stimuli such as hands, faces, or objects



SENSATION & PERCEPTION 4e, Figure 4.6

Lesions to dorsal stream areas cause...

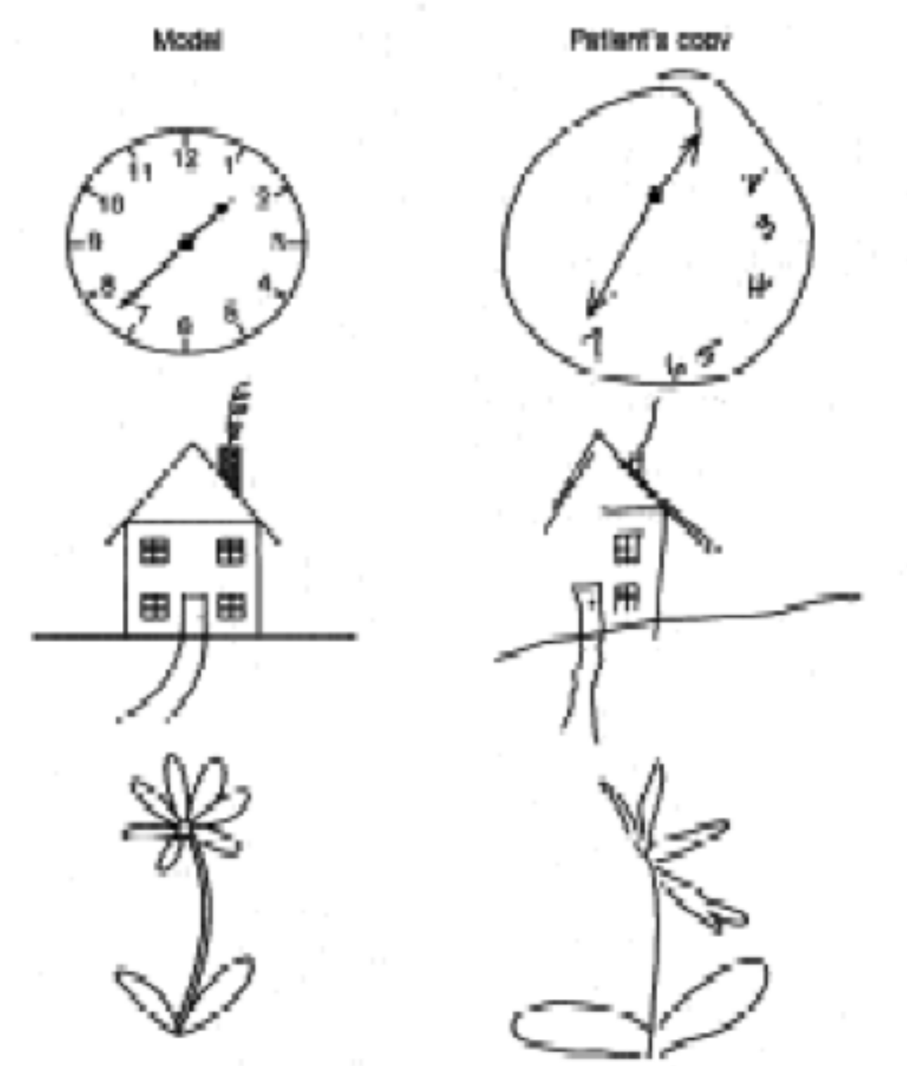
Optic ataxia: inability to visually guide arm movements

Hemispatial neglect: patient is unaware of the contralateral region of space.

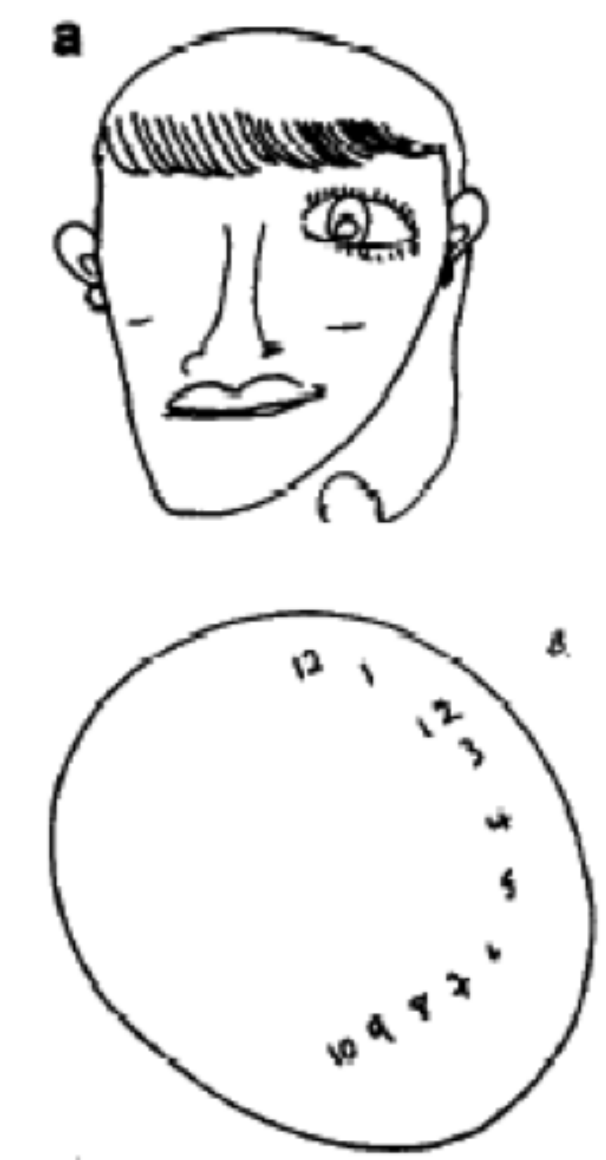
Akinetopsia: inability to perceive motion

Figure 5.27 For the patient with motion blindness, the world appears as if viewed through a strobe light. Rather than see the liquid rise continuously in the teacup, the patient reports seeing the liquid jump from one level to the next.

Copying:



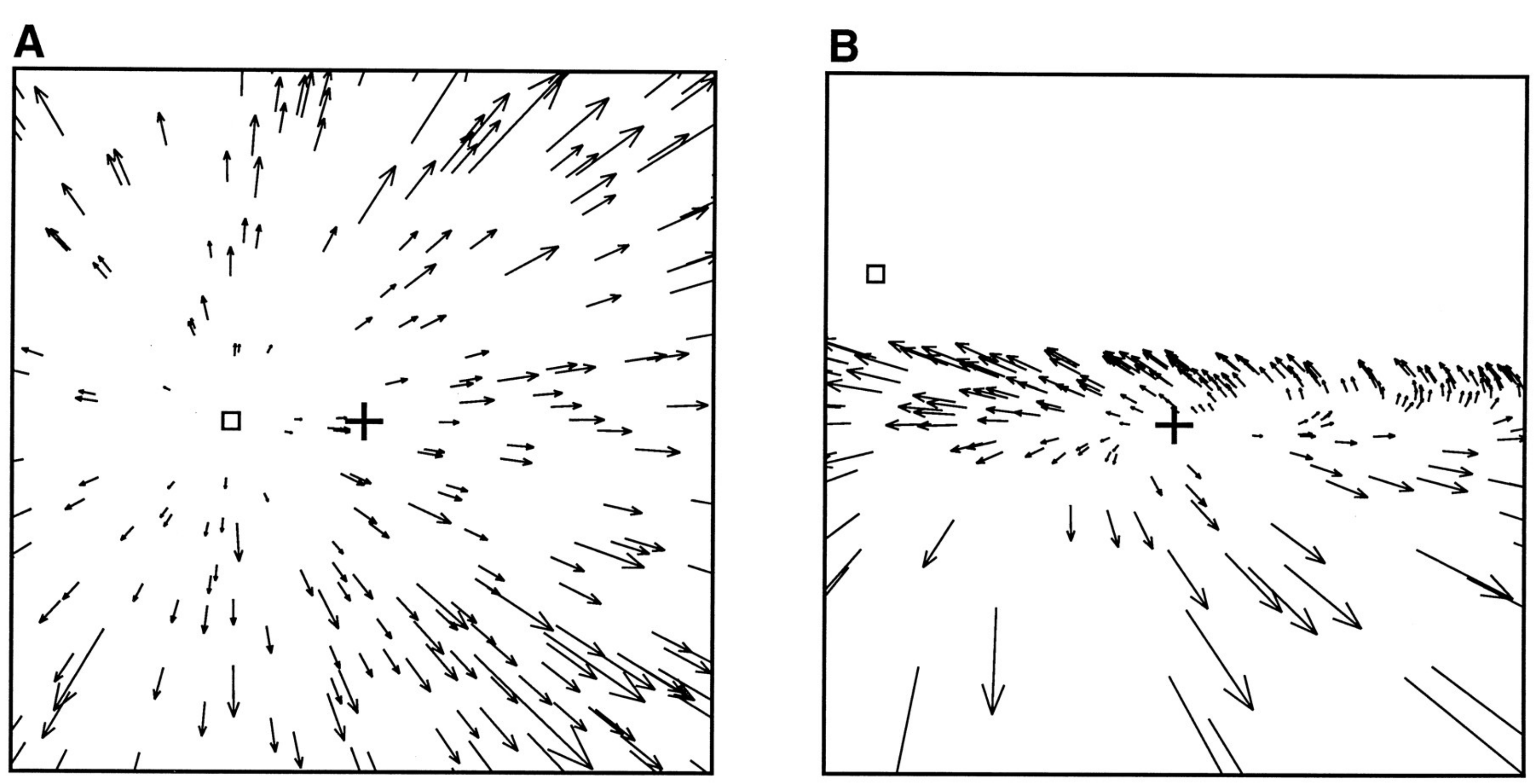
Spontaneous drawing:

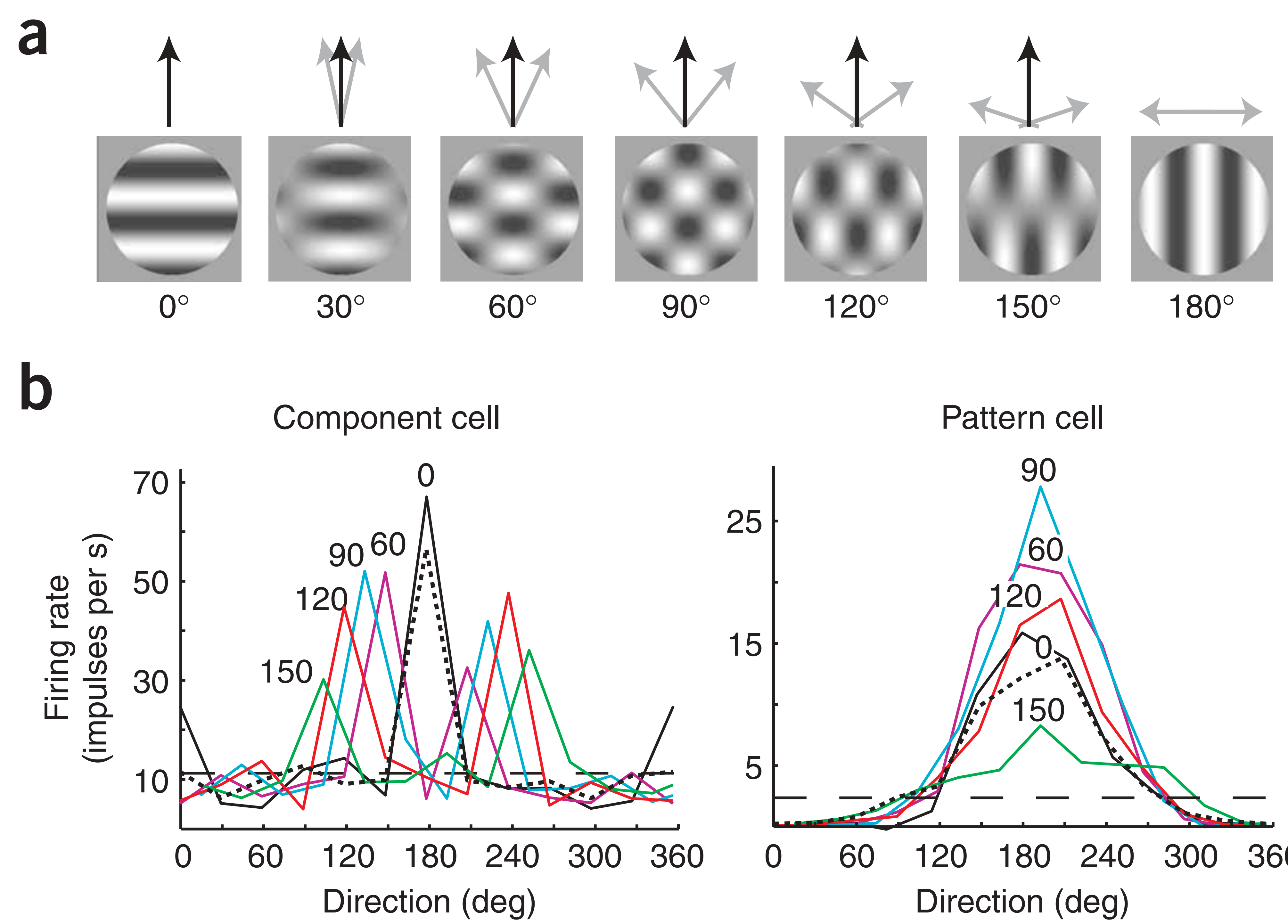


Receptive fields of MT neurons

Neurons in the middle temporal (MT) area respond selectively to motion directions.

Neurons in MST (medial superior temporal) area respond to optic flow patterns





Rust, N.C., Mante, V., Simoncelli, E. P., & Movshon, J. A. (2006). How MT cells analyze the motion of visual patterns. *Nature Neuroscience*, *9*(11), 1421–1431.

Evidence for the two-streams hypothesis

Double dissociations:

lesions to ventral stream areas damage object and shape perception but tend not to affect motion or spatial perception

whereas lesions to dorsal stream areas tend to impair motion and spatial perception, but leave object recognition intact

Functional specialisation in the visual brain

I argue here that research using functional MRI is beginning to answer this long-standing question with new clarity and precision by indicating that at least a few specific aspects of cognition are implemented in brain regions that are highly specialized for that process alone. Cortical regions have been identified that are specialized not only for basic sensory and motor processes but also for the high-level perceptual analysis of faces, places, bodies, visually presented words, and even for the very abstract cognitive function of thinking about another person's thoughts.

Kanwisher, N. (2010). Functional specificity in the human brain: A window into the functional architecture of the mind. *Proceedings of the National Academy of Sciences*, *107*(25), 11163–11170.

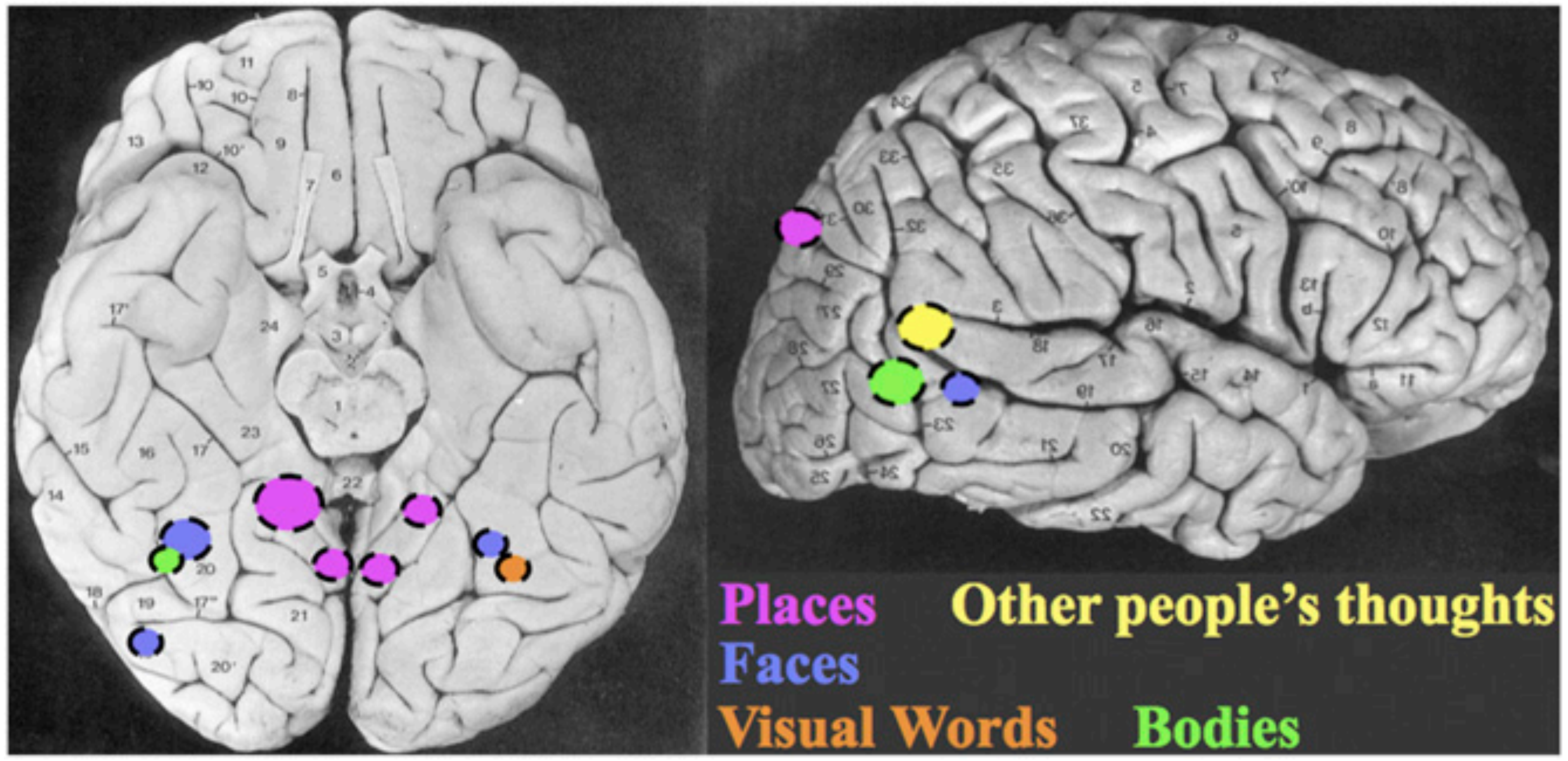


Fig. 1. This schematic diagram indicates the approximate size and location of regions in the human brain that are engaged specifically during perception of faces (blue), places (pink), bodies (green), and visually presented words (orange), as well as a region that is selectively engaged when thinking about another person's thoughts (yellow). Each of these regions can be found in a short functional scan in essentially all normal subjects.

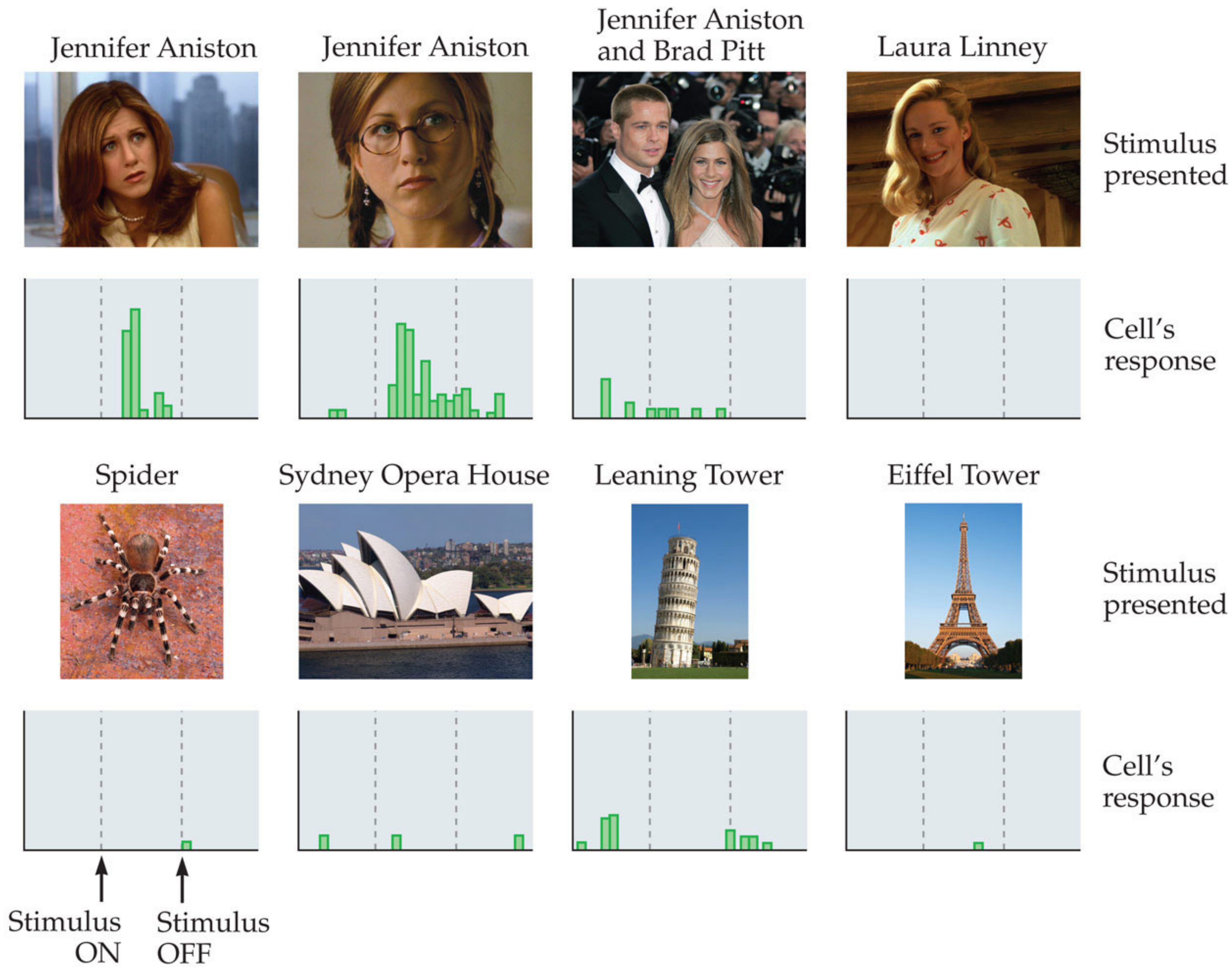
Grandmother cells?

Could a single neuron be responsible for recognizing your grandmother? (Idea originally from Barlow, 1972).

Seems very unlikely for a number of reasons, not least that a system relying on a single neuron to perform a computation would not be very robust.

But … perhaps. At least some people—not me!—believe that this has recently become a more realistic idea: Quiroga et al. (2005) identified a cell that responds specifically to Jennifer Aniston.

Results of recording the activity of one cell in the temporal lobe of a human patient



SENSATION & PERCEPTION 4e, **Figure 4.7**

Does it respond *specifically* to Jennifer Aniston? What about "actors I like" or "women I fancy" or "part of favourite TV series" or whatever—almost infinite possible search space before being able to conclude the specificity to Jennifer Aniston.

Face perception and the FFA: evidence for causal relationship to perception

https://dx.doi.org/10.1523%2FJNEUROSCI.2609-12.2012

Parvizi, J., Jacques, C., Foster, B. L., Withoft, N., Rangarajan, V., Weiner, K. S., & Grill-Spector, K. (2012). Electrical Stimulation of Human Fusiform Face-Selective Regions Distorts Face Perception. *Journal of Neuroscience*, *32*(43), 14915–14920.

Neuroscience of object recognition: Caveats and outstanding questions

1. Most terms ill-defined, e.g. "parallel processing", "feedback" (Douglas & Martin, 1991)

2. Receptive versus Projective Fields (Lehky & Sejnowski, 1988): what is the *functional role* rather than the *best stimulus* of a neuron?

> Examination of the receptive fields of individual units does not make apparent what the network is doing, and interpretations other than that of extracting curvatures from shaded images are likely to spring to mind. While this model network obviously does not establish that receptive fields in the cortex which resemble those developed by the network are engaged in shading analysis, it does raise questions about conventional interpretations of the functions of receptive fields, not only in visual pathways, but in other sensory systems as well. Understanding the function of a neuron within a network appears to require not only knowledge of the pattern of input connections forming its receptive field, but also knowledge of the pattern of output connections, which forms its projective field. Indeed, the same neuron may have a number of different functions if it projects to several regions.

Neuroscience of object recognition: Caveats and outstanding questions

3. Anatomy: cortical pathways are *not* based on *functional* anatomy, i.e. know about number of fibres only, not their importance or "power":

The lateral geniculate nucleus is the best understood thalamic relay and serves as a model for all thalamic relays. Only 5–10% of the input to geniculate relay cells derives from the retina, which is the driving input. The rest is modulatory and derives from local inhibitory inputs, descending inputs from layer 6 of the visual cortex, and ascending inputs from the brainstem. These modulatory inputs control many features of retinogeniculate transmission. … By contrast, the other main thalamic relay of visual information, the pulvinar region, is largely a higher-order relay, since much of it relays information from layer 5 of one cortical area to another. All thalamic relays receive a layer-6 modulatory input from cortex, but higher-order relays in addition receive a layer-5 driver input. Corticocortical processing may involve these corticothalamocortical 're-entry' routes to a far greater extent than previously appreciated. If so, the thalamus sits at an indispensable position for the modulation of messages involved in corticocortical processing.
(Sherman & Guillery, 2002)

4. How much have we even really understood about V1? "Silent neurons", sampling bias, size of cells …

The End

Felix Wichmann

Neural Information Processing Group
Eberhard Karls Universität Tübingen