Human Perception

CS 294-10: Virtual Reality & Immersive Computing
EECS, UC Berkeley
Fall 2017

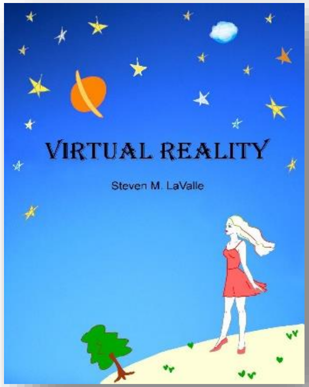
Achin Bhowmik, PhD

CTO & EVP of Engineering
Starkey Hearing Technologies
Berkeley, CA and Eden Prairie, MN

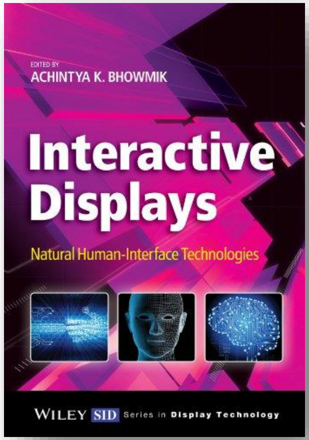
Introduction

- Current role (brand new!)
  - CTO & EVP of Engineering at Starkey Hearing Technologies
- Intel (Oct 2000 – July 2017)
  - VP & GM of Perceptual Computing Group (3D/depth-sensing, immersive computing, computer vision and deep learning chips, drones and robots, VR and MR)
  - Chief of Staff of Personal Computing Group
  - Display and Video/Image processing, mobile computer architecture R&D
- Teach at various universities
  - Berkeley, Stanford, KSU
  - IIT Gandhinagar, KHU Seoul
- Community
  - Fellow, Executive Committee and Board member of SID
  - Chair, SID Augmented & Virtual Reality Track
  - Board of Directors, OpenCV
  - Board of Advisors, Fung Institute at UC Berkeley

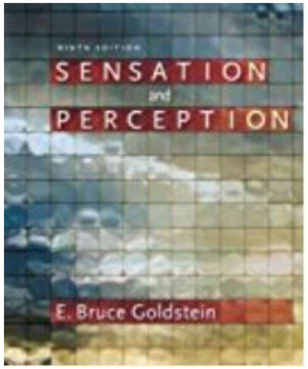
Books (Optional Reading)



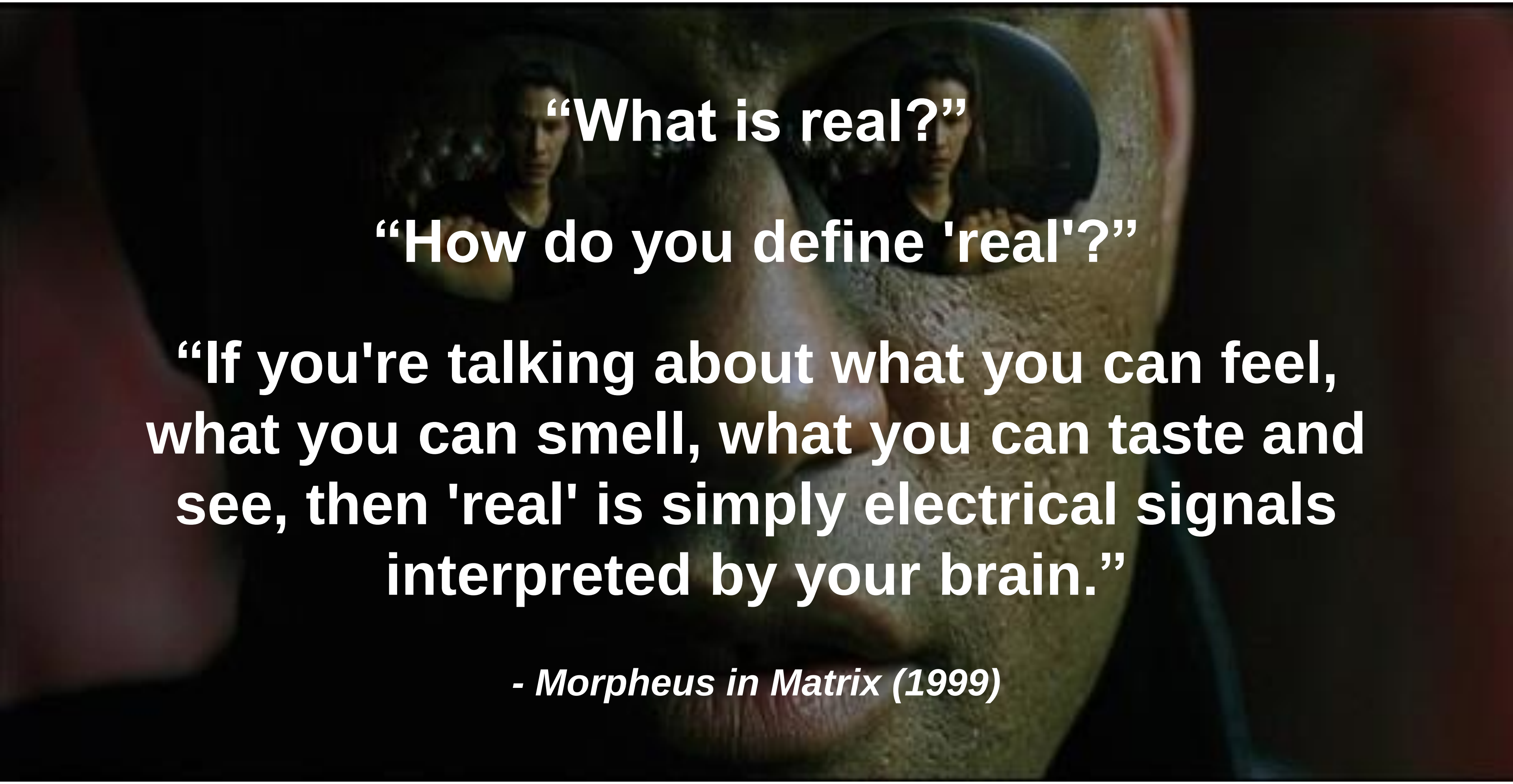
Virtual Reality, Steven M. LaValle, to be published by Cambridge University Press
http://vr.cs.uiuc.edu/



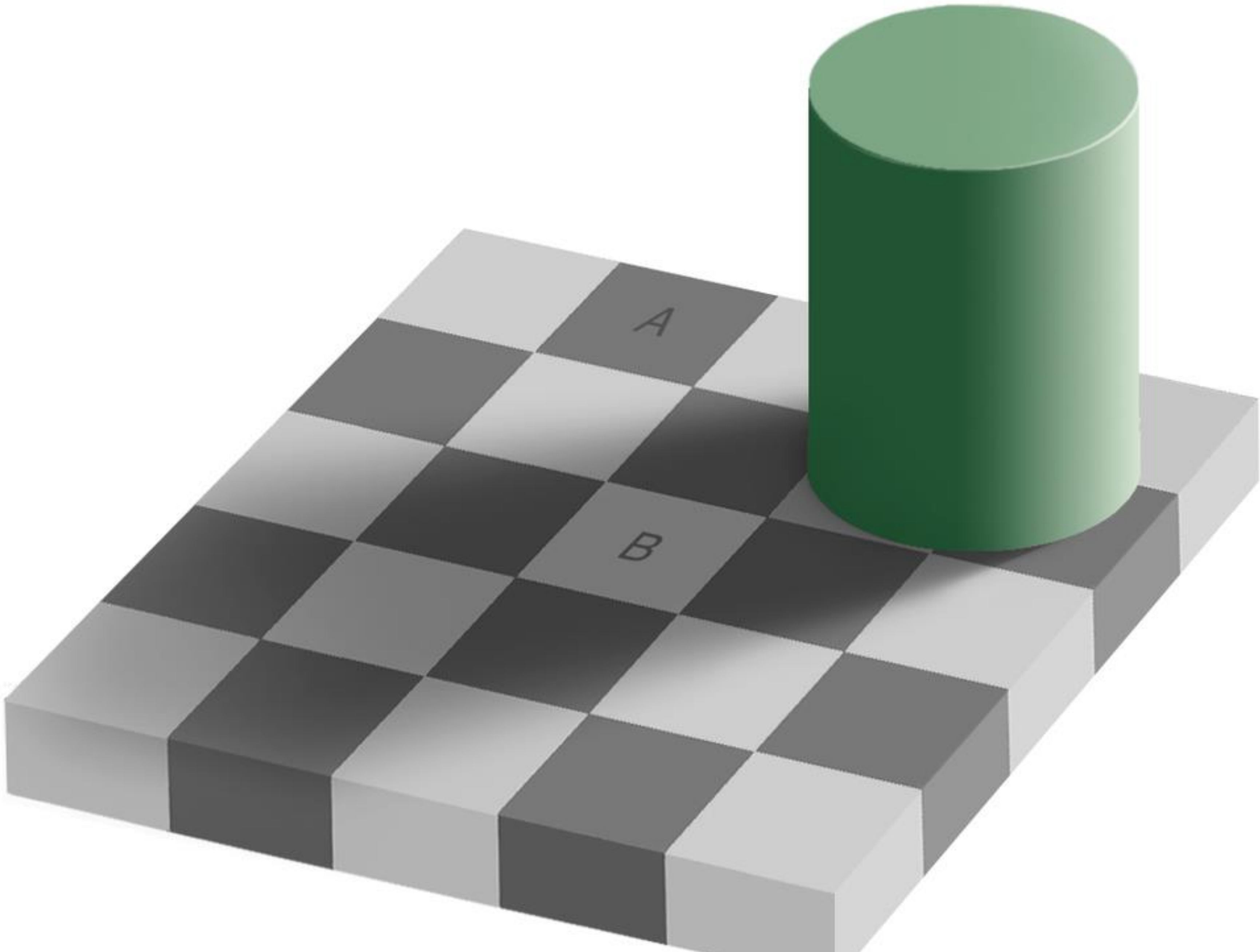
Interactive Displays: Natural Human-Interface Technologies, Achintya K. Bhowmik, Wiley (2014)
http://www.wiley.com/WileyCDA/WileyTitle/productCd-1118631374.html

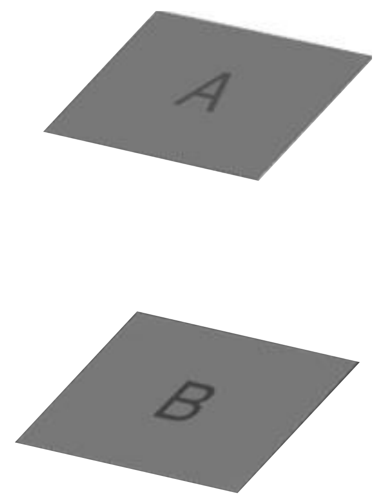


Sensation and Perception, E. Bruce Goldstein, Wadsworth Publishing (2013)
https://www.amazon.com/Sensation-Perception-Bruce-Goldstein-ebook/dp/B00BF3VMSA

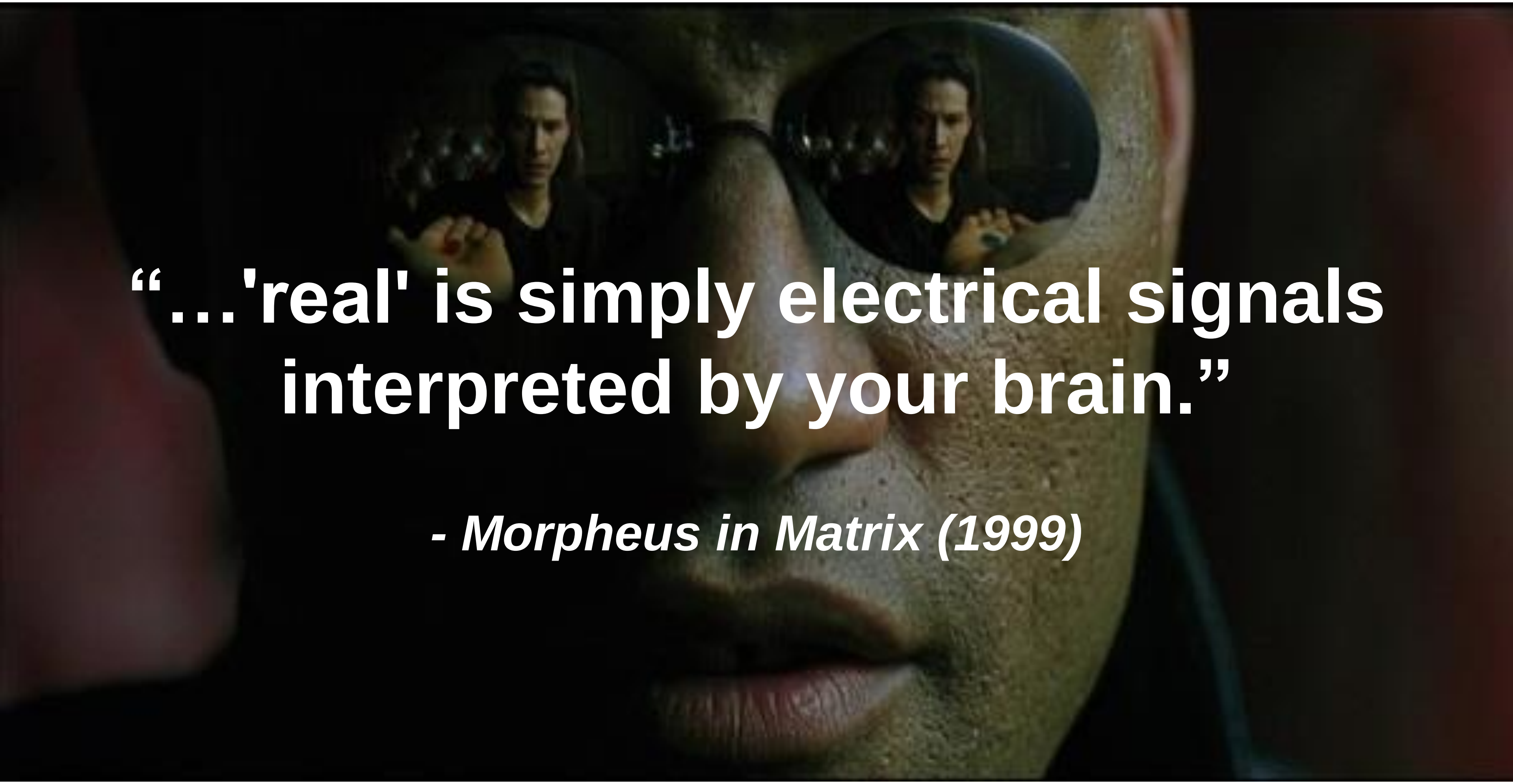

“What is real?”
“How do you define 'real'?”
“If you're talking about what you can feel, what you can smell, what you can taste and see, then 'real' is simply electrical signals interpreted by your brain.”
- Morpheus in Matrix (1999)

www.youtube.com/watch?v=WlEzvdlYRes

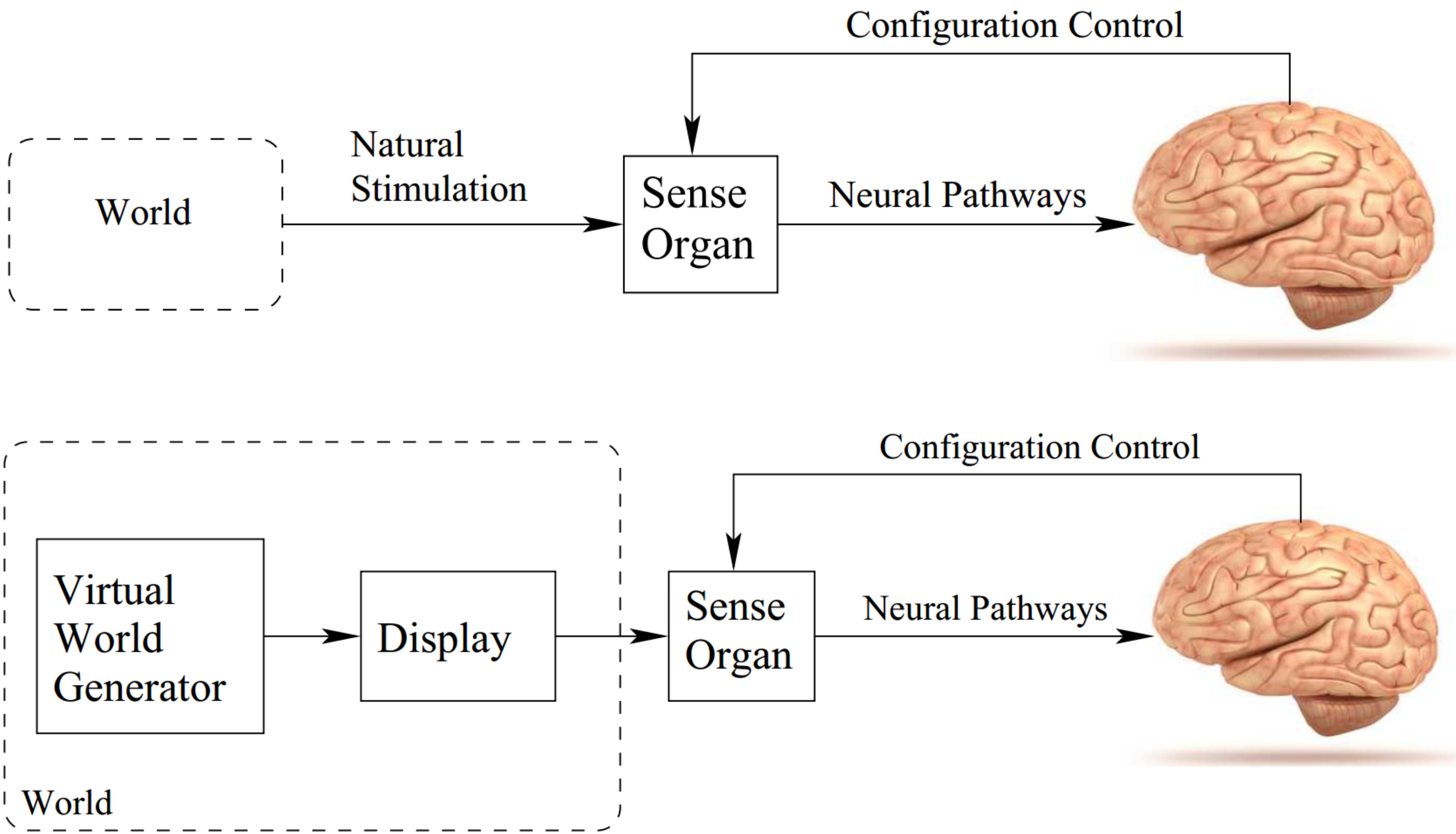

A
B


A
B

www.youtube.com/watch?v=aFPtc8BVdJk

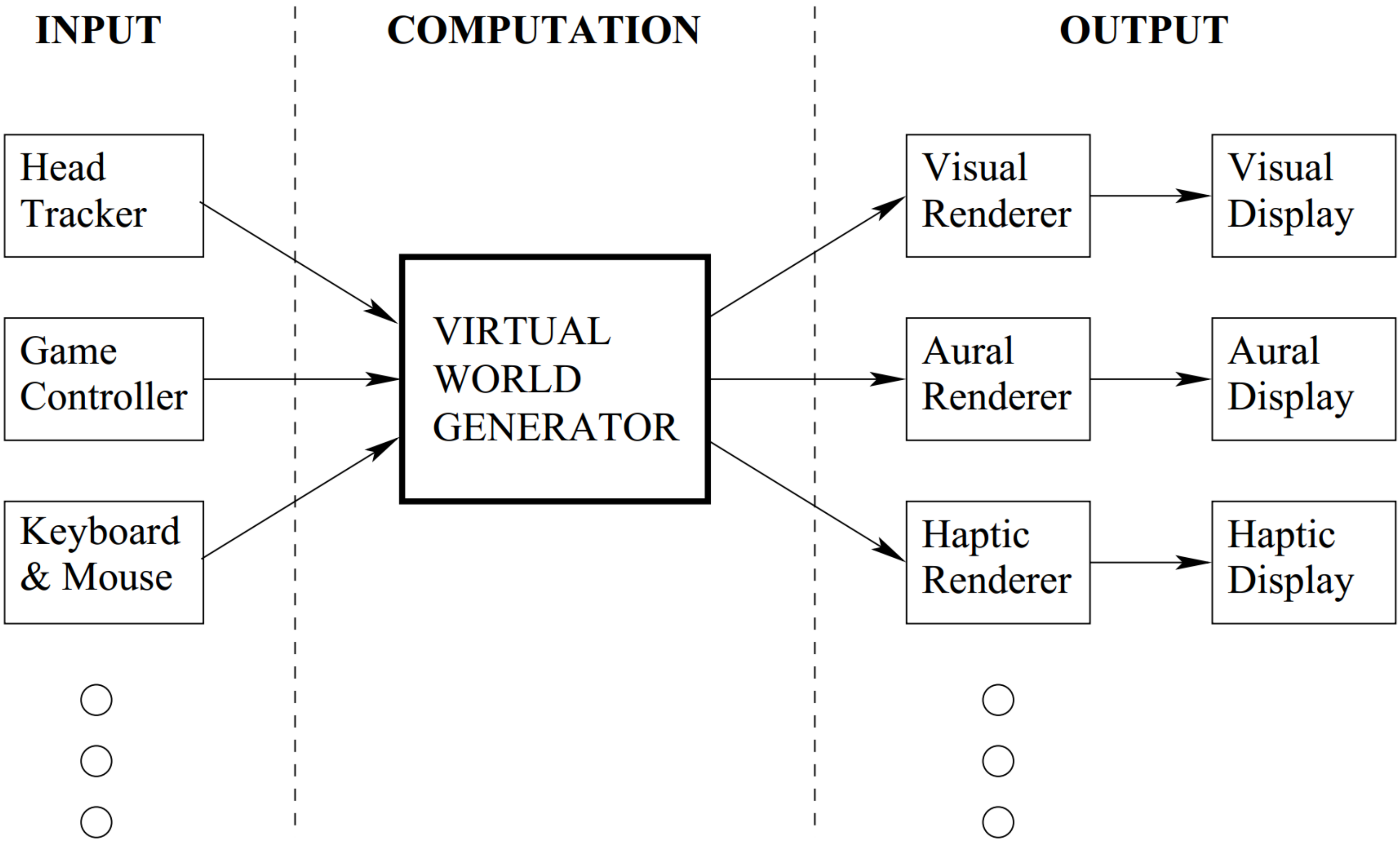

“…'real' is simply electrical signals interpreted by your brain.”
- Morpheus in Matrix (1999)

The Real and the Virtual Worlds



S. Lavalle

System at a Glance…

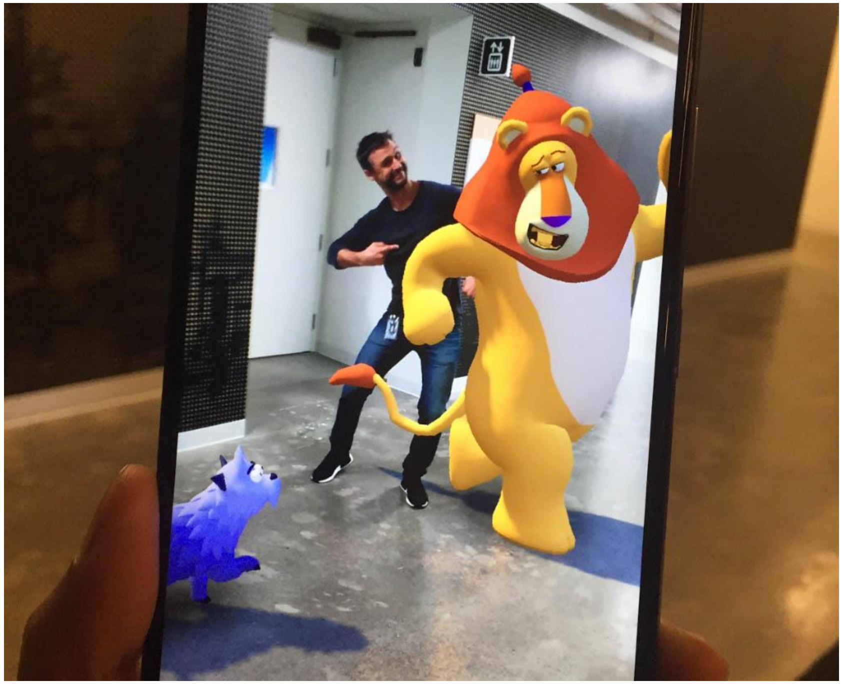


S. Lavalle

Virtual Reality Experience

Courtesy: Michael Abrash (now at Facebook Oculus)

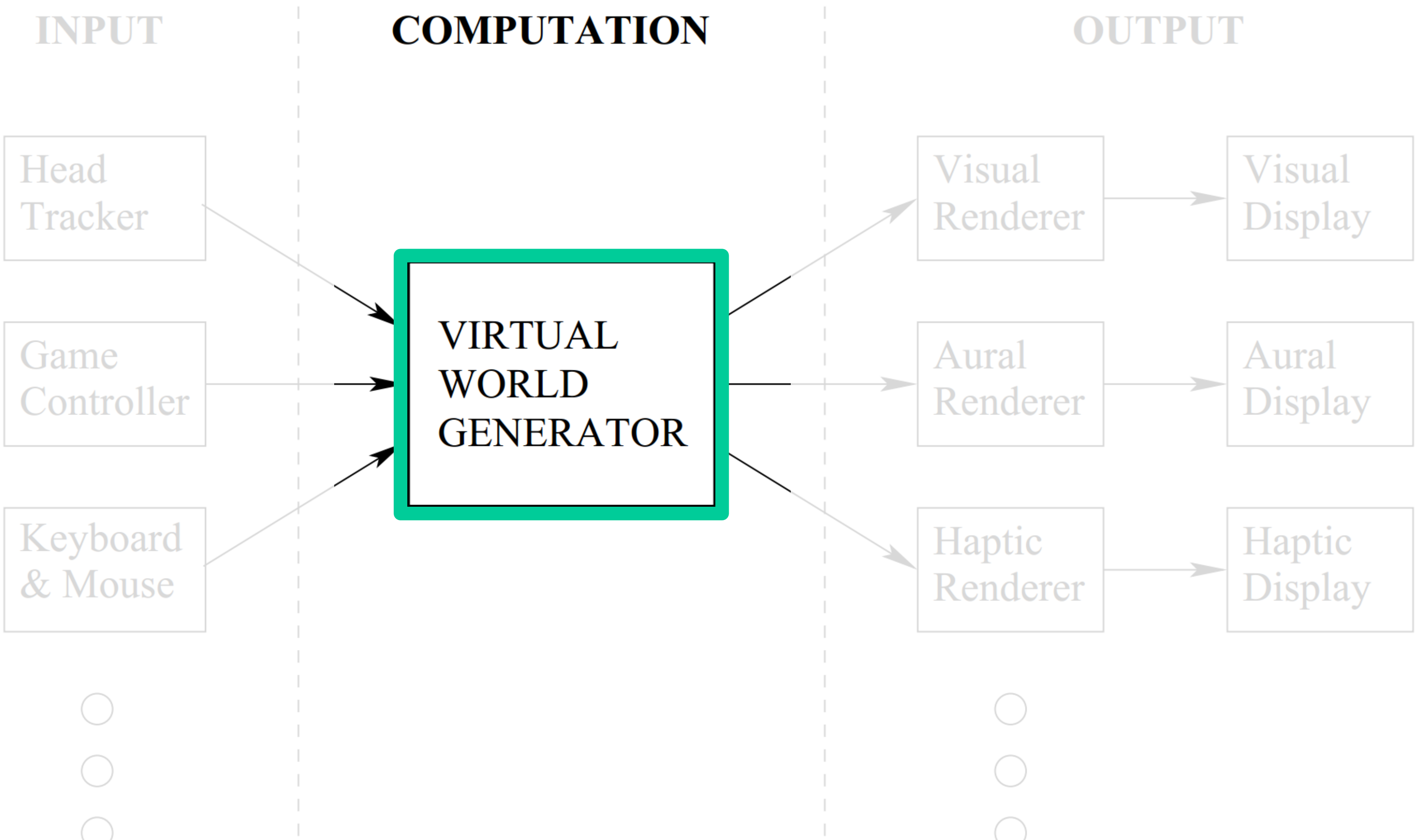
Augmented & Mixed Reality Experiences



Google Tango

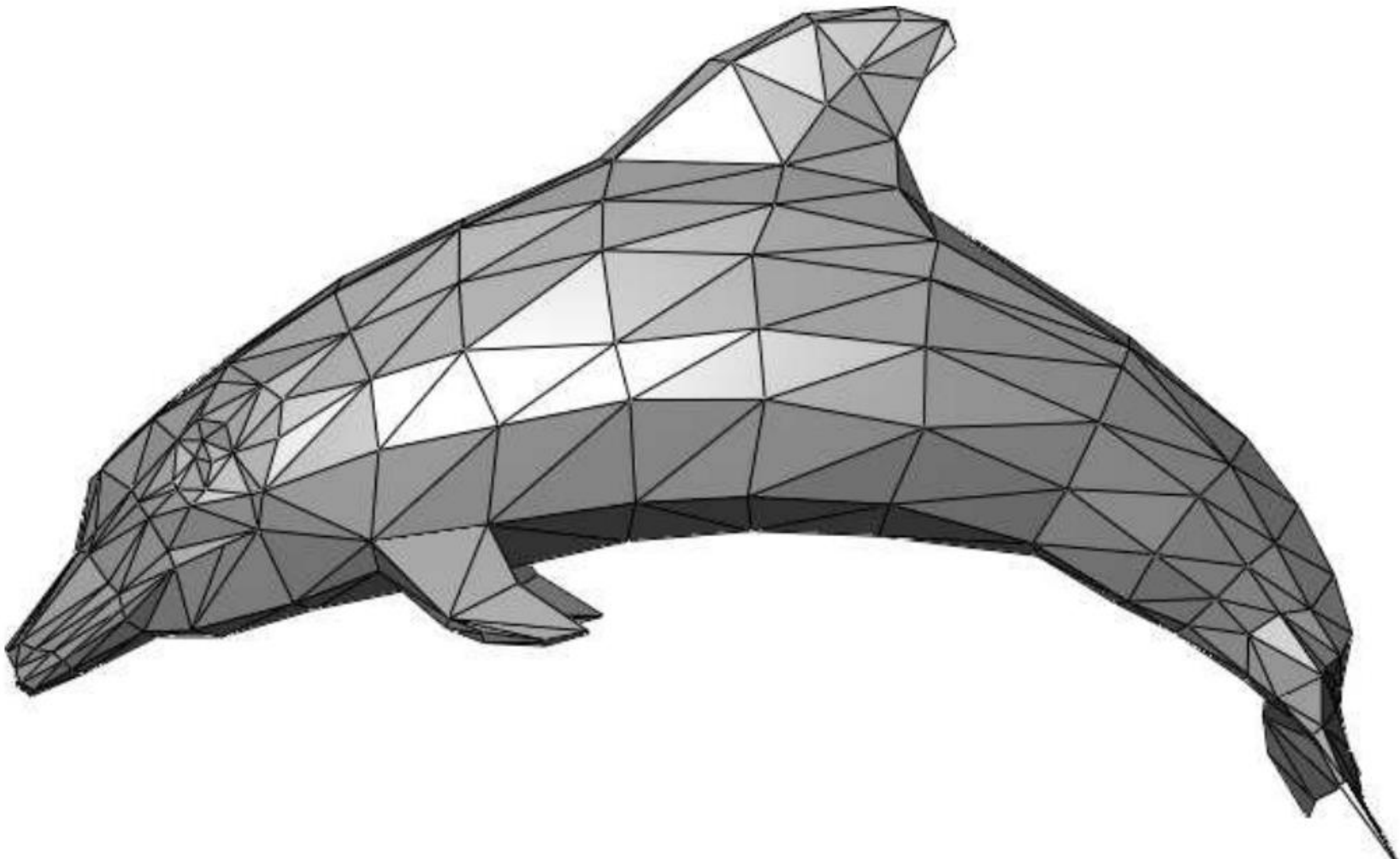
Microsoft Hololens

System at a Glance…



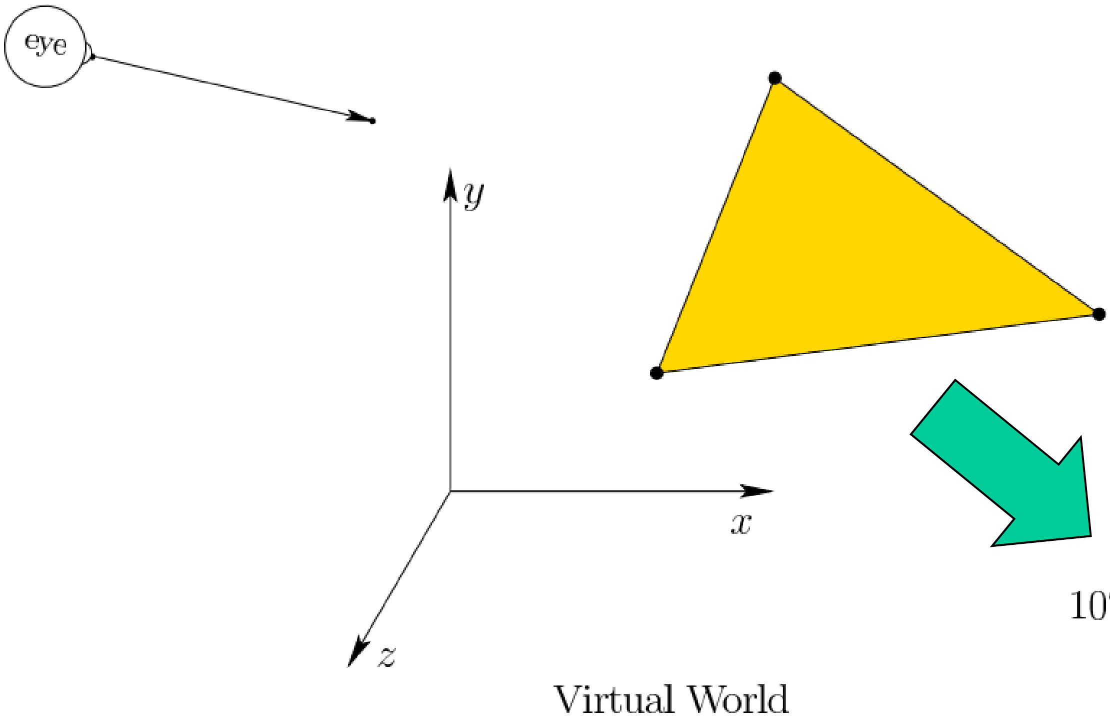
S. Lavalle

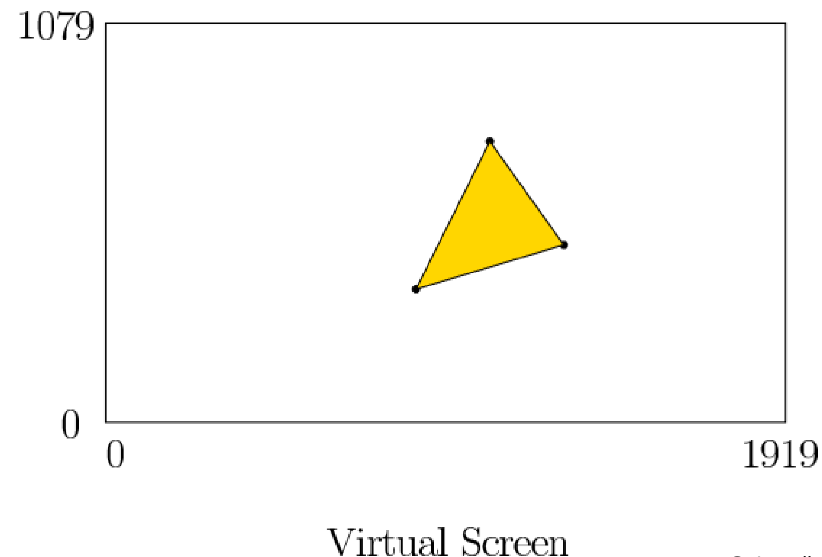
Geometric Models in VR/AR/MR: Mesh of 3D Primitives



Chrschn

3D Primitives and Viewing Transformations



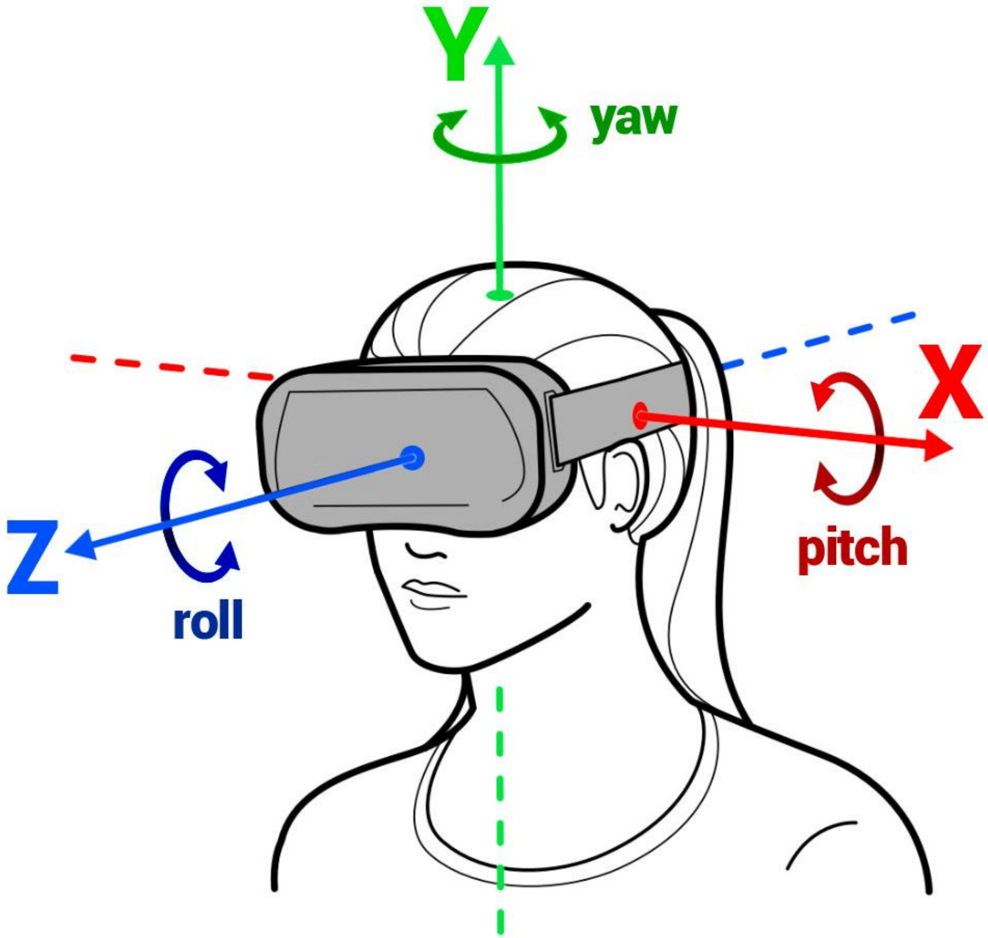


Transformation matrices:

$$T = T_{vp} T_{can} T_{eye} T_{rb}$$

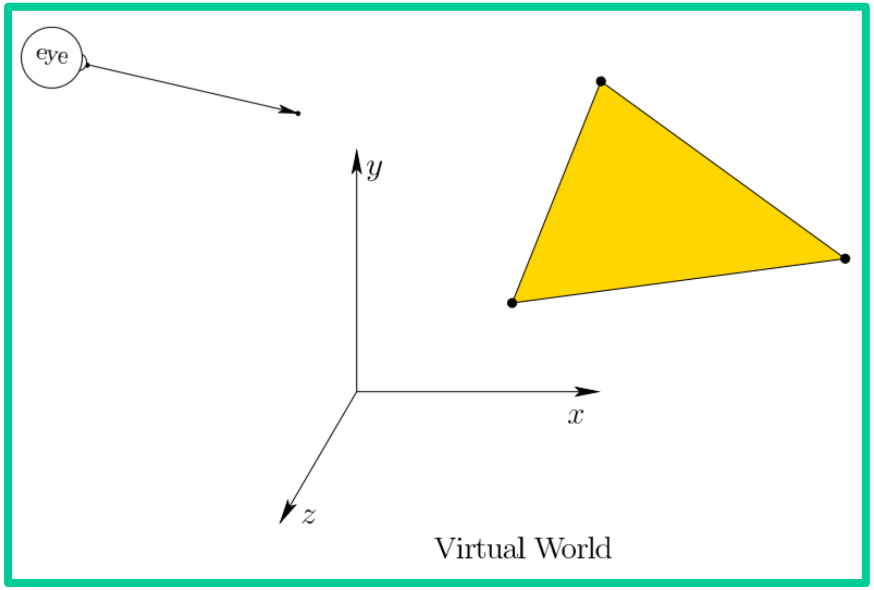
S. Lavalle

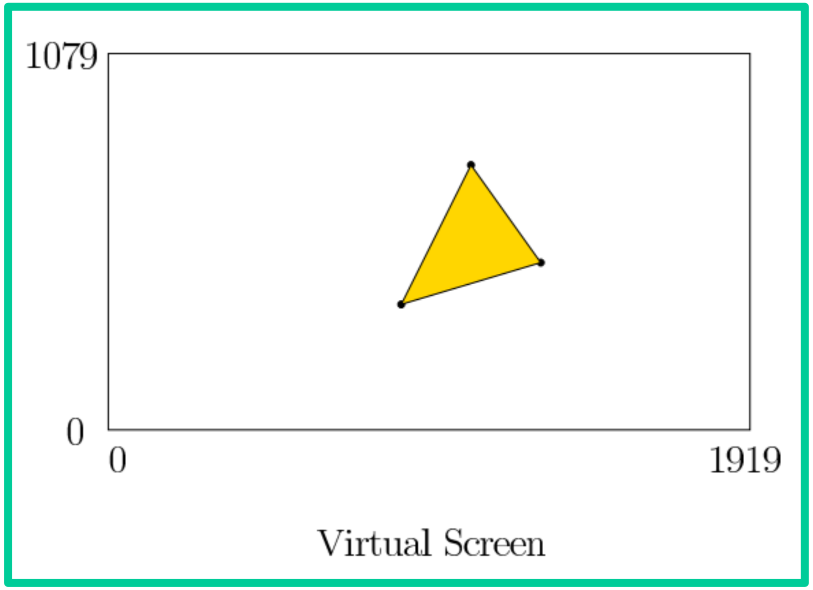
3D Rotation and 3D Translation: Six Degrees of Freedom

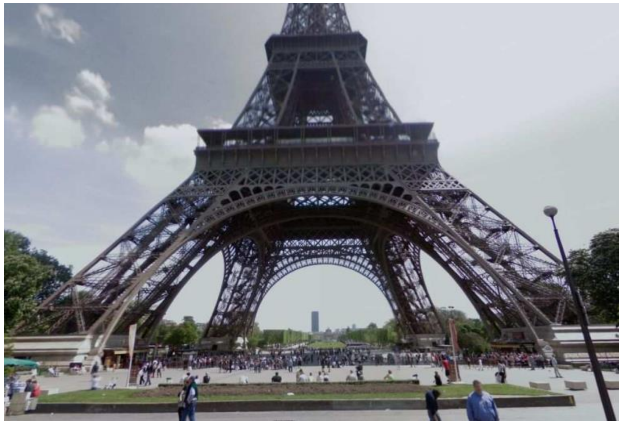


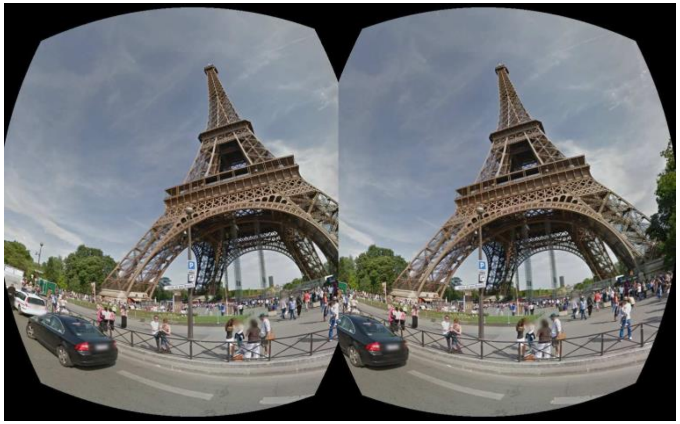
http://dsky9.com/rift/vr-tech-6dof/

After applying all the viewing transformations…





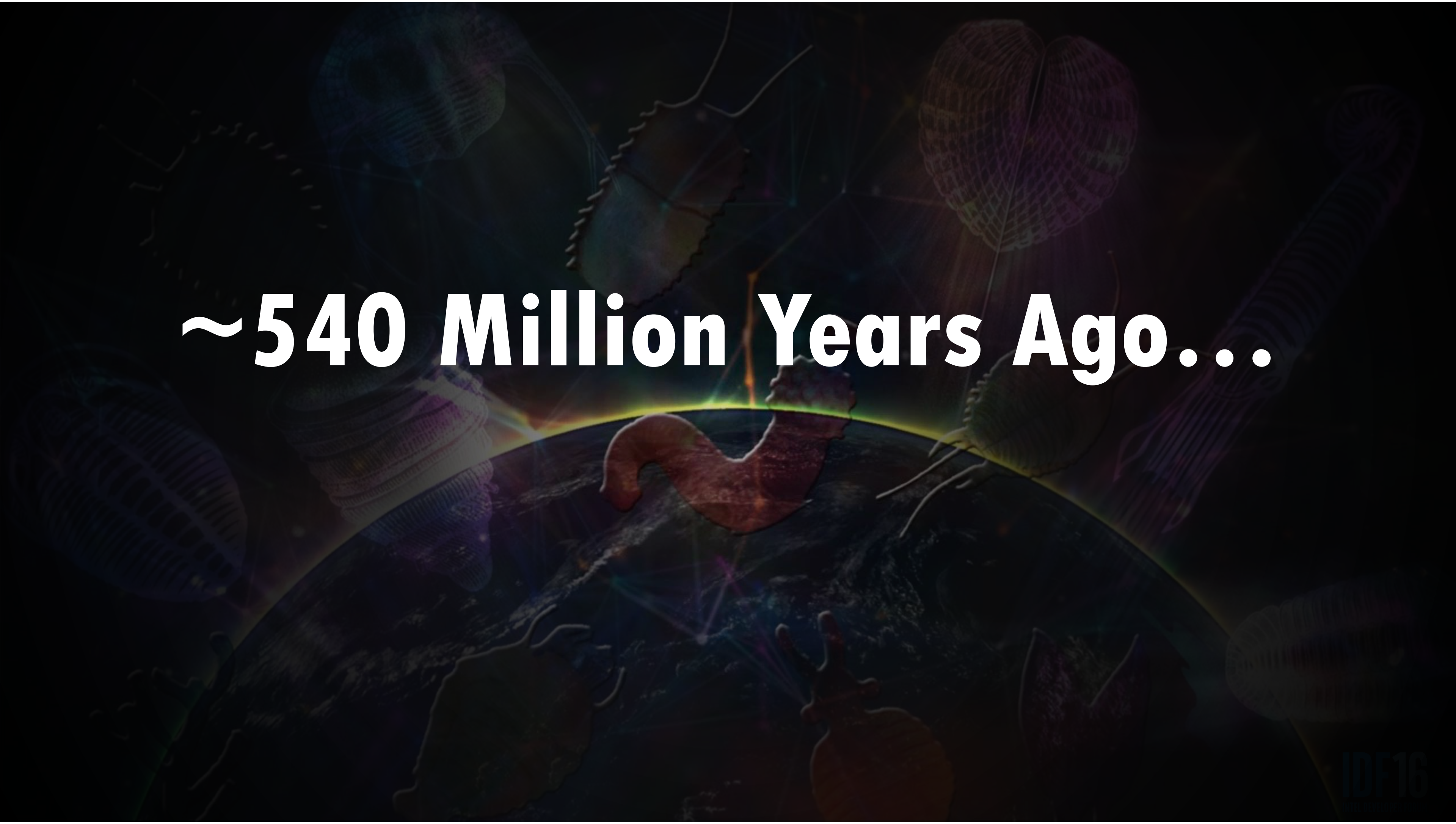




(In)famous “User Experience” Issues

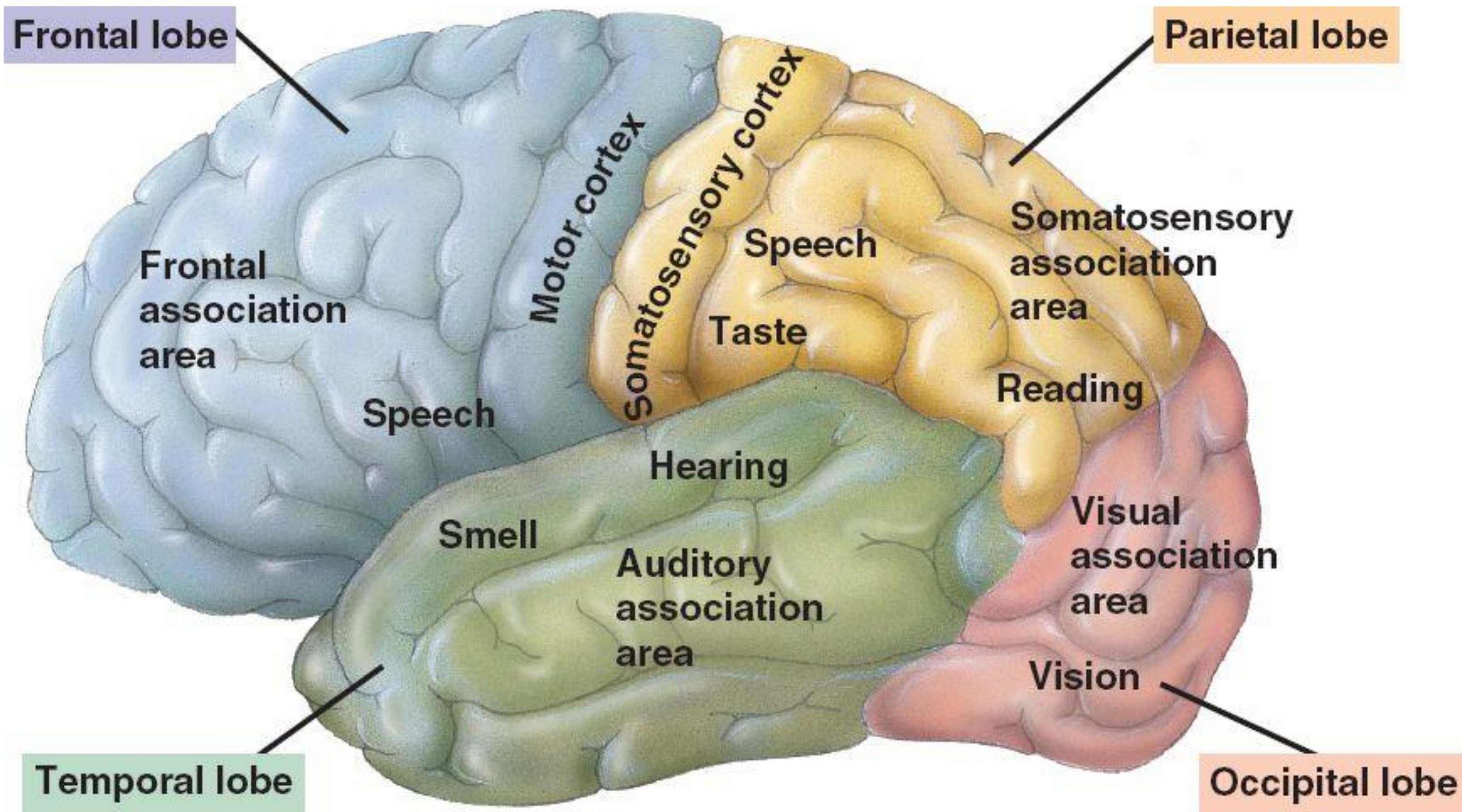
- Inconsistent cues from the visual and vestibular sensory systems
  - Motion that is “seen but not “felt”
  - Motion that is “felt” but not “seen”
  - Both systems detect motion but they do not correspond
- Visual experiences
  - “Screen-door” effects due to “low” display resolutions
  - Field-of-view (FOV) limitations
  - Motion artifacts due to “low” display frame-rates
- System/ergonomic issues
  - Weight, comfort, tether, eye relief, …
  - Power dissipation, battery life, …
  - Tracking setup vs. “inside-out” tracking, 3-DOF vs. 6-DOF
- Inconsistent oculomotor cues
  - Eye accommodation/convergence mismatch
  - Incorrect focus/blur cues
- Missing or inconsistent proprioception cues
- Other issues we will inevitably discover in the future…

A Look at Biology…


~540 Million Years Ago…

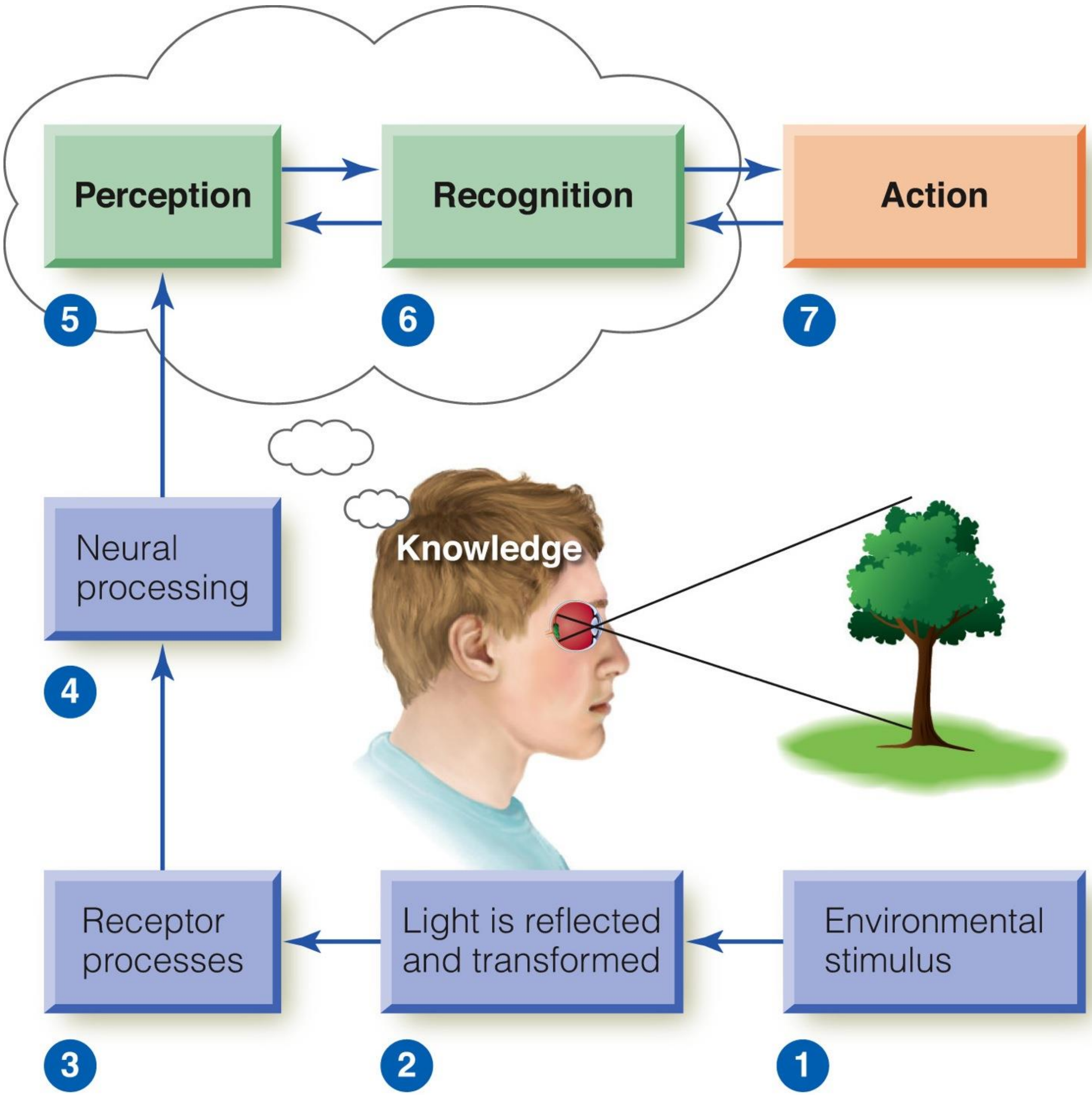

The Cambri
Explosion

Human Cerebral Cortex:
Sensory Processing Modules

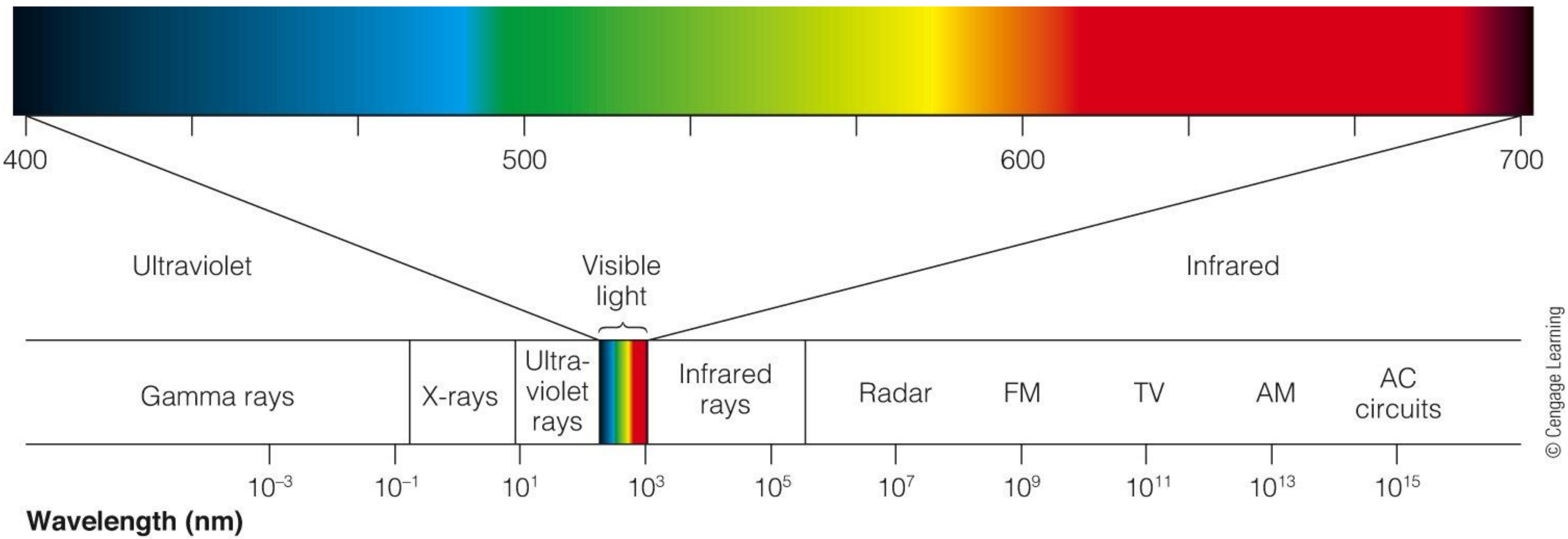


http://bio1152.nicerweb.com/Locked/media/ch48/cerebral.html

Human Visual Perception

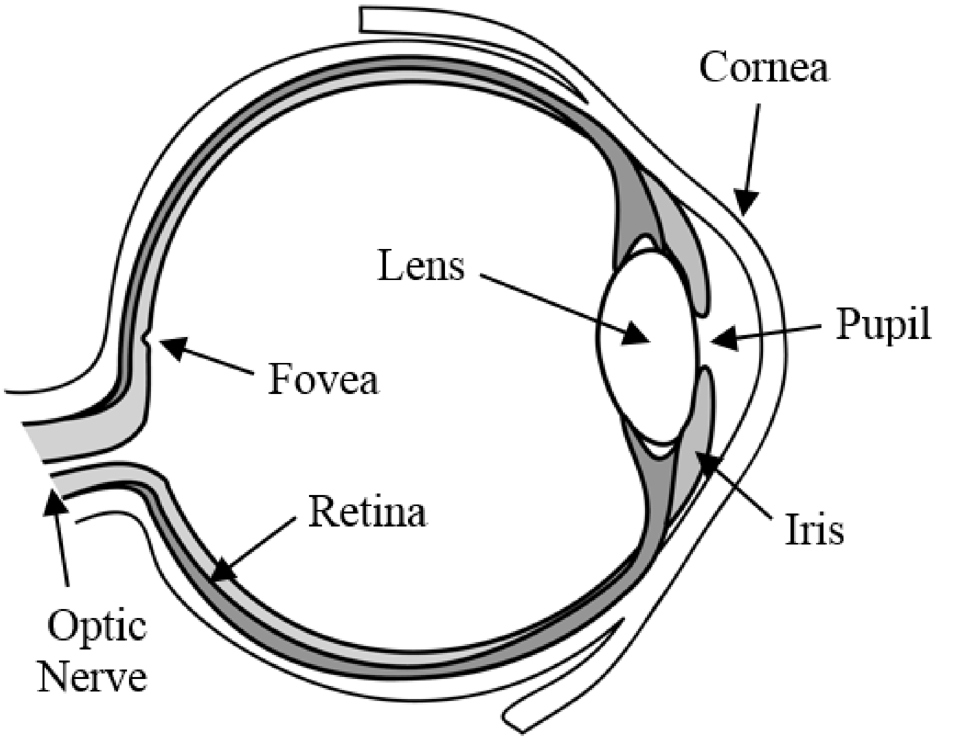

Perception
Recognition
Action
5
6
7
Neural processing
4
Knowledge
Receptor processes
Light is reflected and transformed
Environmental stimulus
3
2
1

Light: the Stimulus for Vision



- Electromagnetic spectrum
  - Visible spectrum for humans ranges from ~400 to ~700nm.

The Eye: Transduction Organ for Vision



- Light enters the eye through the pupil.
- Focused by the cornea (80%, fixed) and lens (20%, variable) to a sharp image on the retina.
- Rods and cones are the visual receptors in the retina that contain visual pigment.
- The optic nerve carries information from the retina toward the brain.

The Animal Eye Evolution

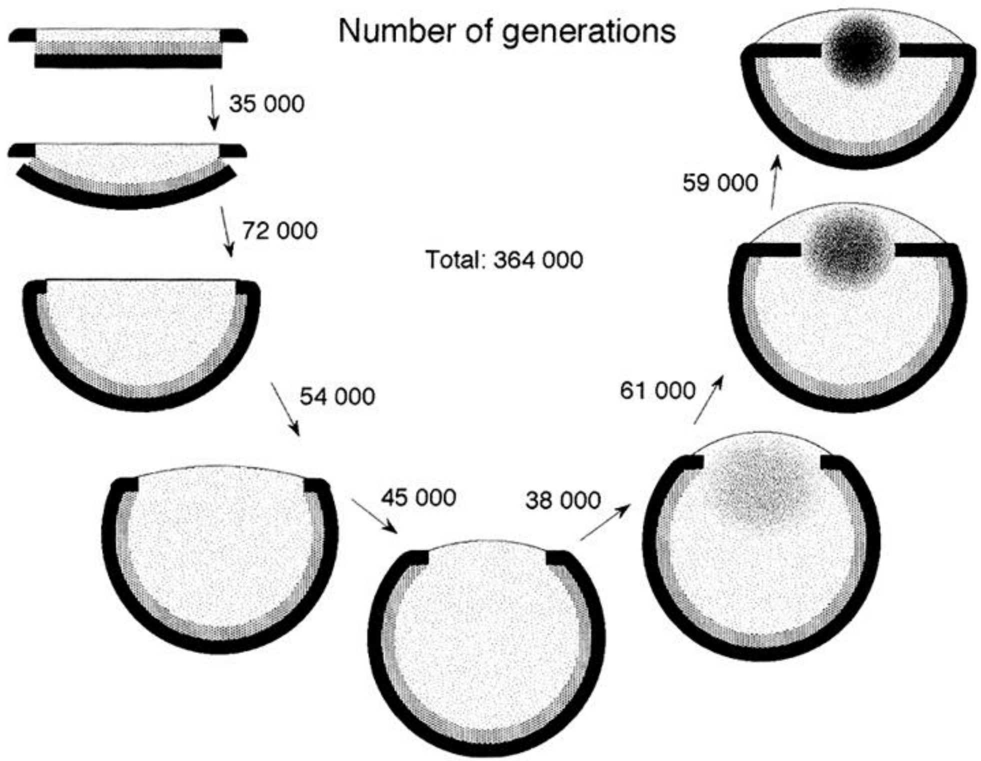
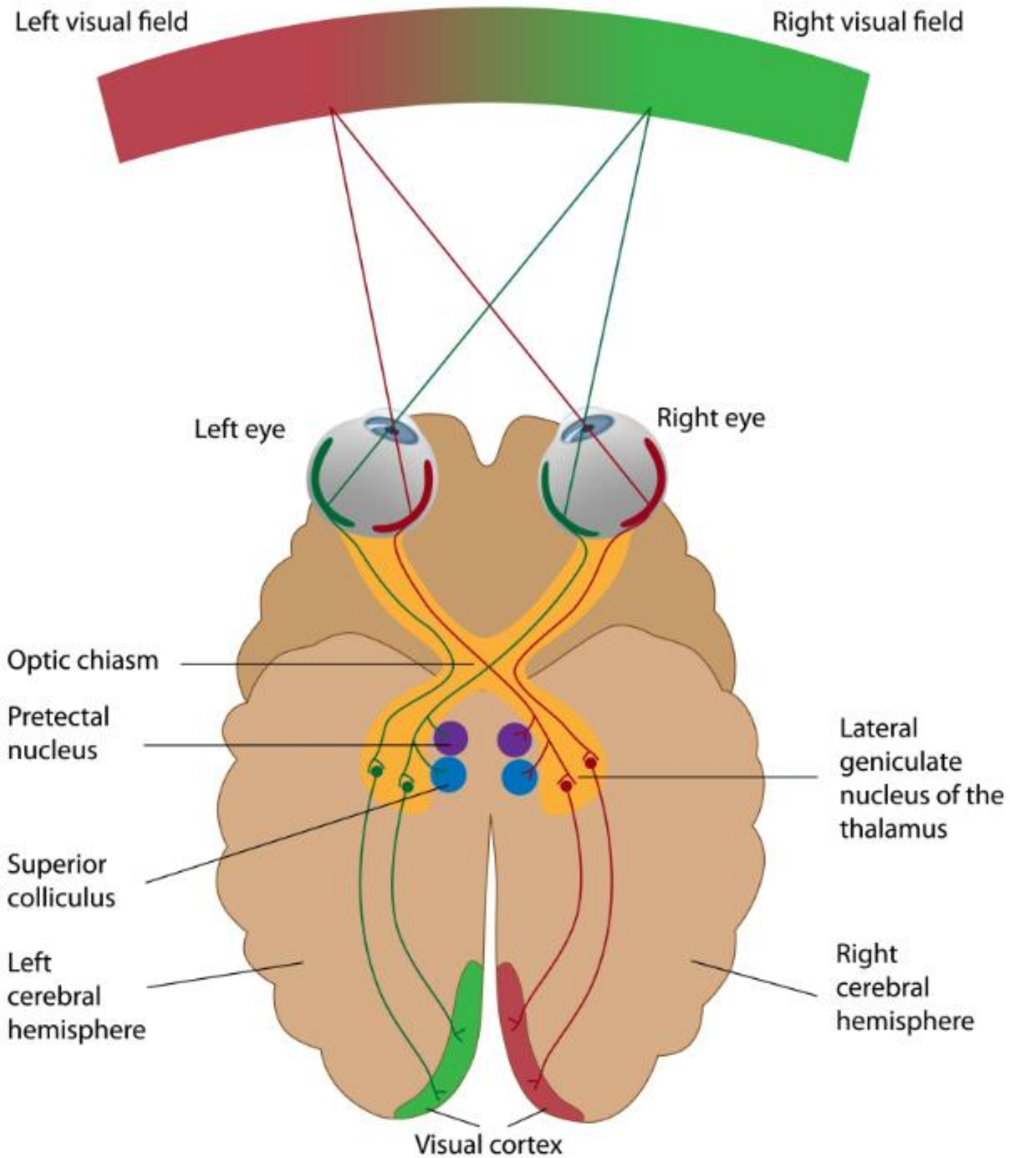


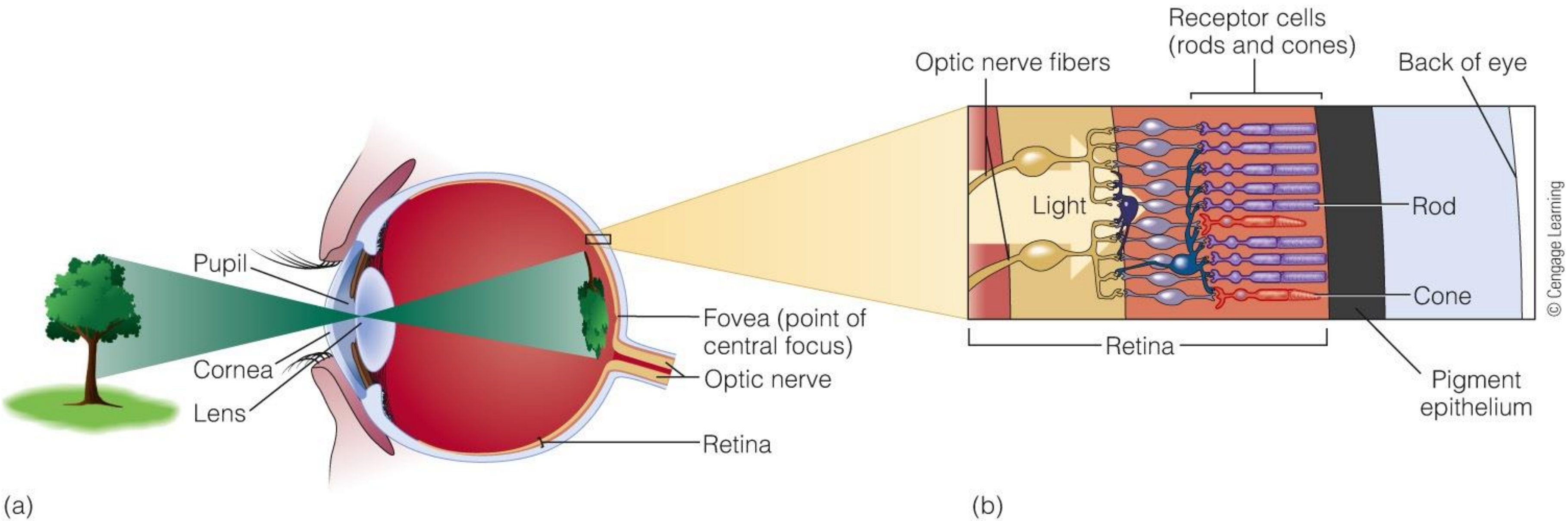
Fig. 1.6 A patch of light sensitive epithelium can be gradually turned into a perfectly focussed camera-type eye if there is a continuous selection for improved spatial vision. A theoretical model based on conservative assumptions about selection pressure and the amount of variation in natural populations suggest that the whole sequence can be accomplished amazingly fast, in less than 400 000 generations. The number of generations is also given between each of the consecutive intermediates that are drawn in the figure. The starting point is a flat piece of epithelium with an outer protective layer, an intermediate layer of receptor cells, and a bottom layer of pigment cells. The first half of the sequence is the formation of a pigment cup eye. When this principle cannot be improved any further, a lens gradually evolves. Modified from Nilsson and Pelger (1994).

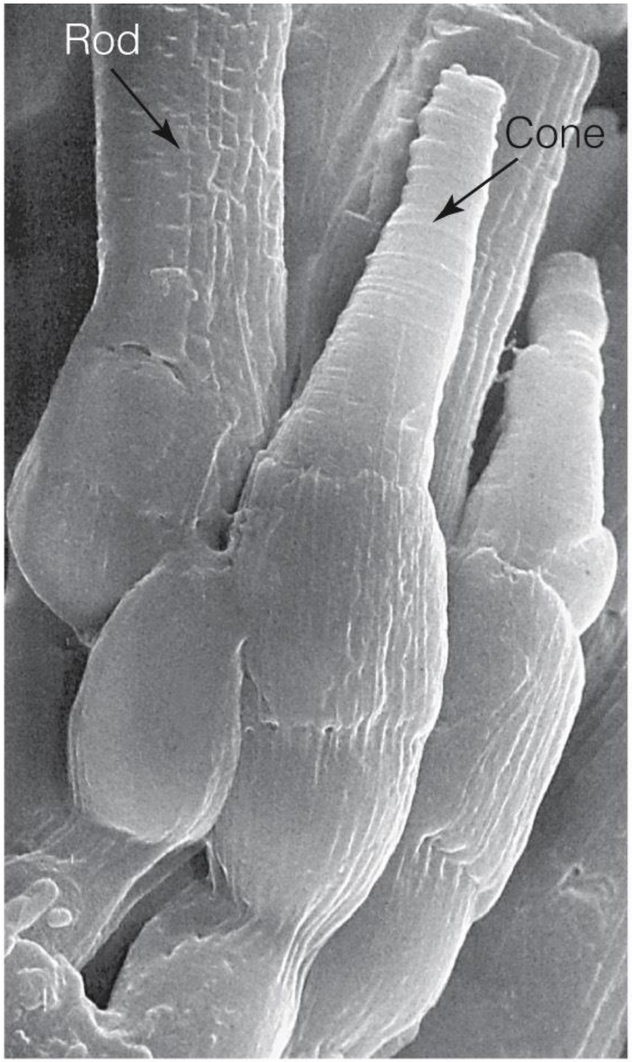
Animal Eyes. Land & Nilsson. Oxford Univ. Press

Binocular Vision System

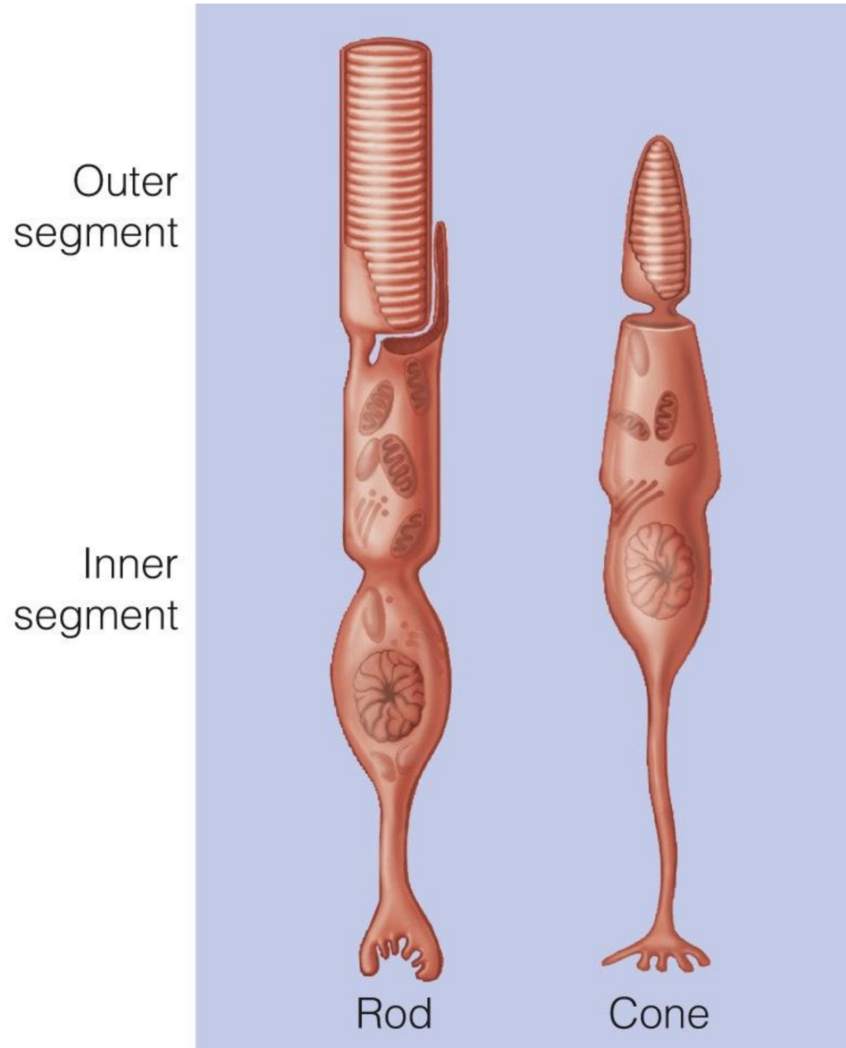


Transforming of Light Energy Into Electrical Energy





(a)

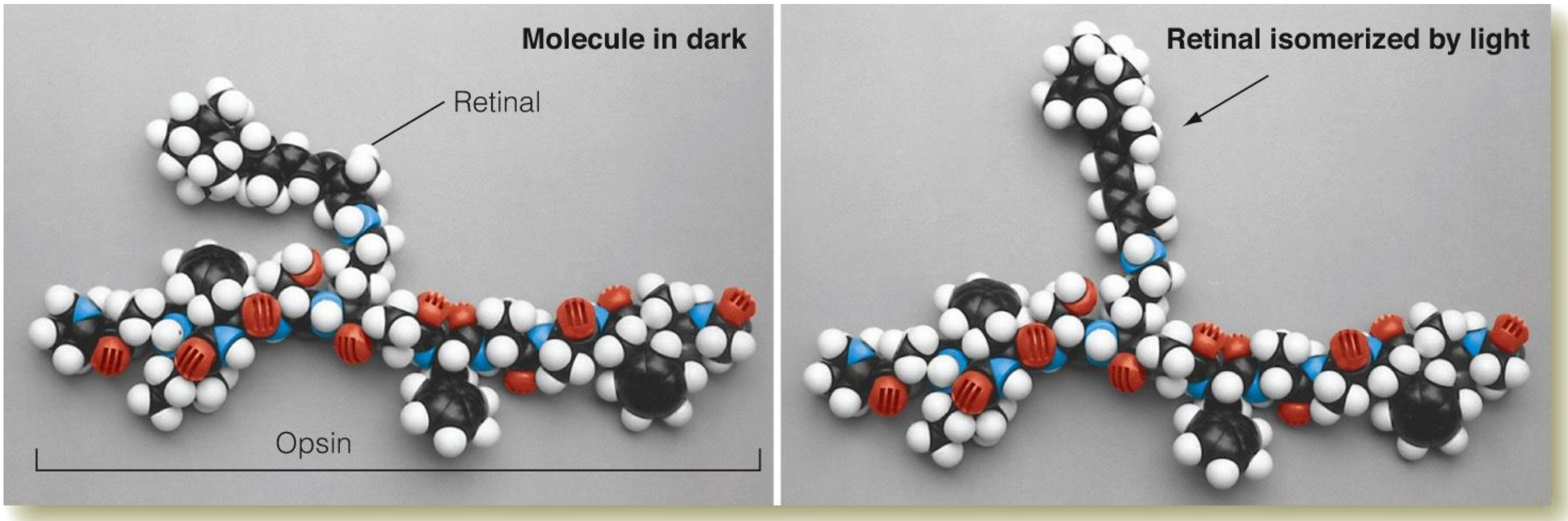


(b)

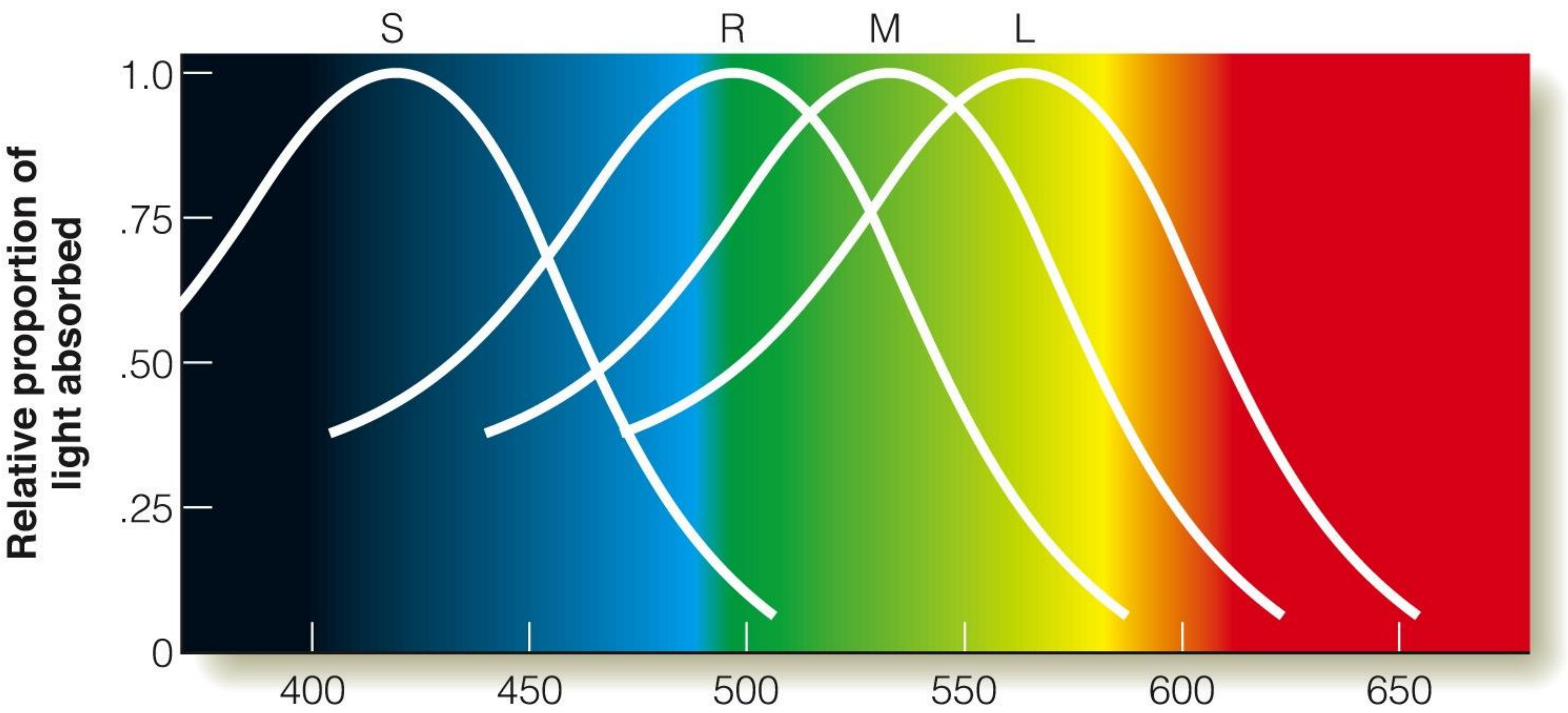
Transforming of Light Energy Into Electrical Energy

- Receptor outer segments contain Visual pigment molecules, which have two components:
    - Opsin - a large protein
    - Retinal - a light sensitive molecule

- Visual transduction occurs when the retinal absorbs one photon
    - Retinal changes it shape, called isomerization


Molecule in dark
Retinal
Opsin
Retinal isomerized by light

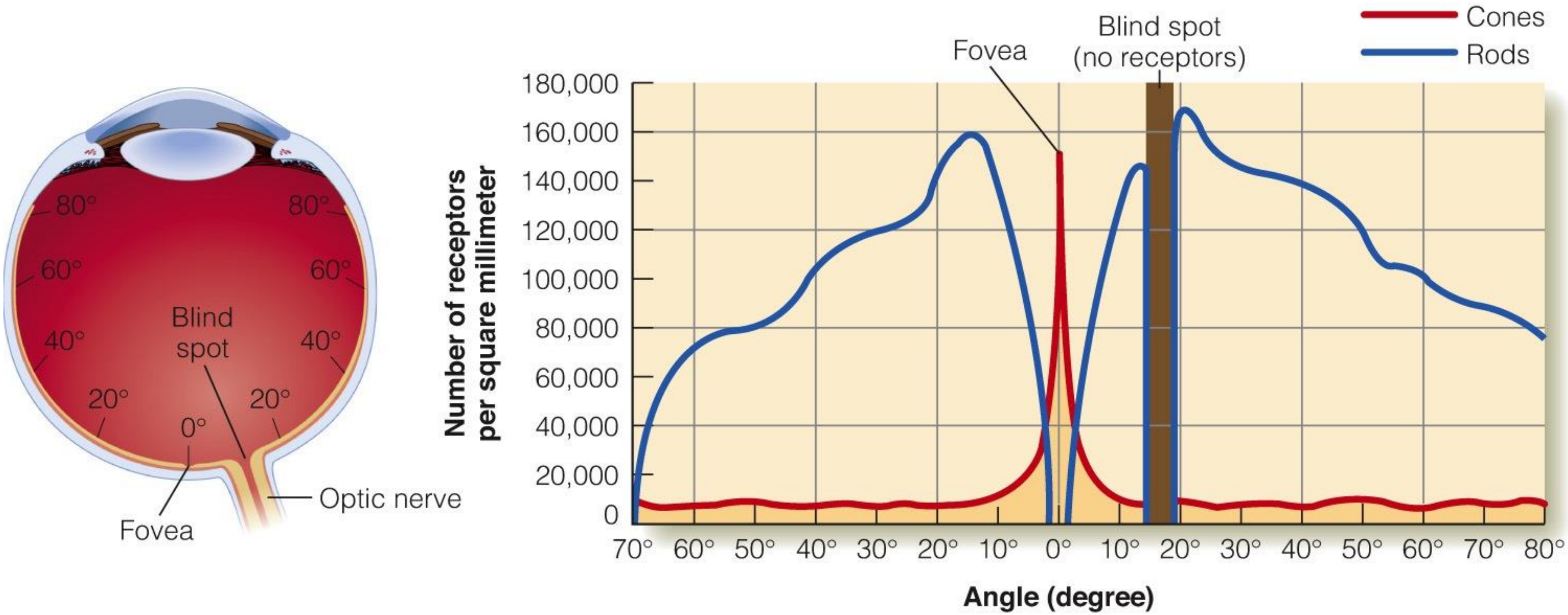
(a) (b)



Based on Dartnall, H. J. A.,
Bowmaker, J. K., & Mollon, J. D. (1983). Human visual pigments: Microspectrophotometric results from the eyes of seven persons. *Proceedings of the Royal Society of London B, 220*, 115–130. Reprinted by permission.

Rod pigment absorbs best at 500 nm, Cone pigments absorb best at 419nm, 531nm, and 558nm. Absorption of all cones equals the peak of 560nm in the spectral sensitivity curve.

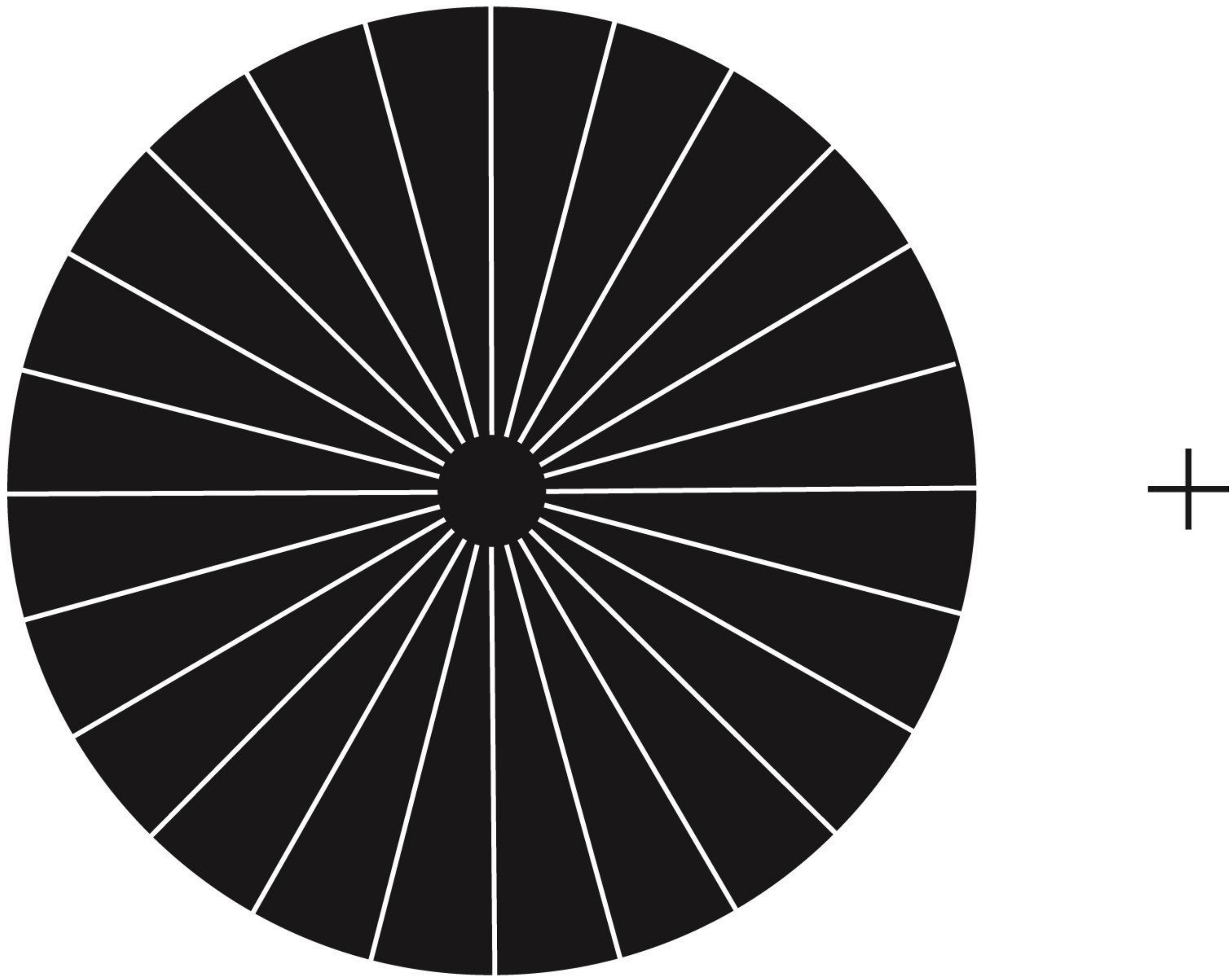
Distribution of Rods and Cones



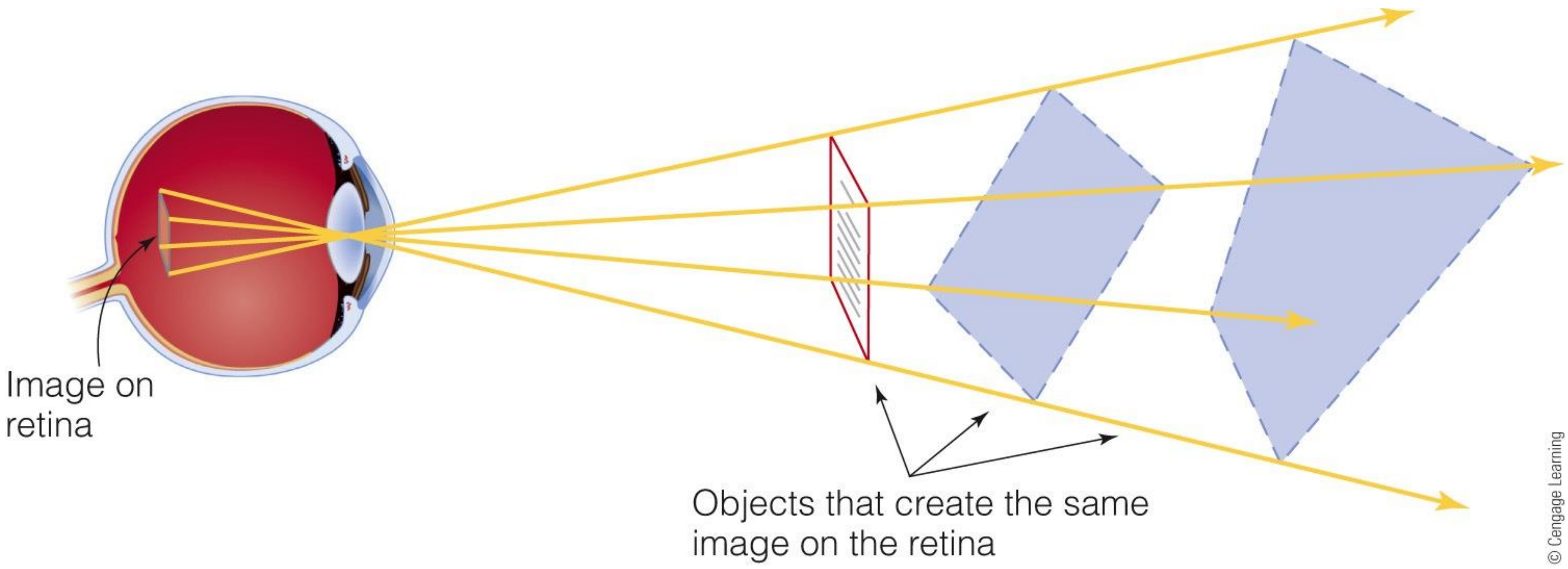
- Fovea consists solely of cones.
- Peripheral retina has both rods and cones.
- More rods than cones in periphery.

http://blog.gingkoapp.com/essays/seeing-an-idea

QWE	RTY	UIO	PAS	DFG	HJK	LZX	CVB



Projection of 3D Scene on 2D Retina




MAKE POVERTY HISTORY
CoolOpticalIllusions.com


CoolOpticalIllusions.com

julianbeever.net

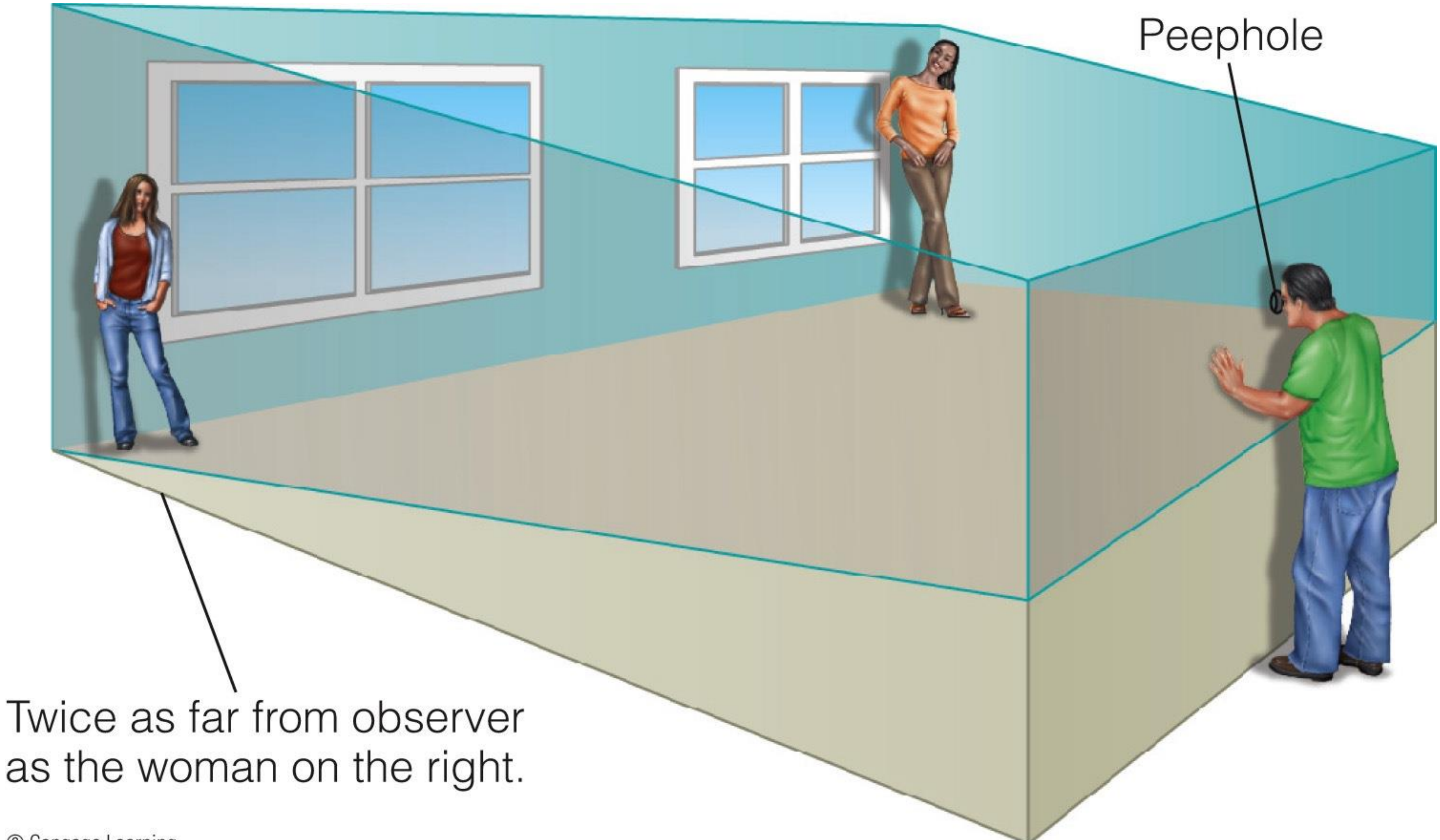
julianbeever.net

julianbeever.net

julianbeever.net

The Ames Room

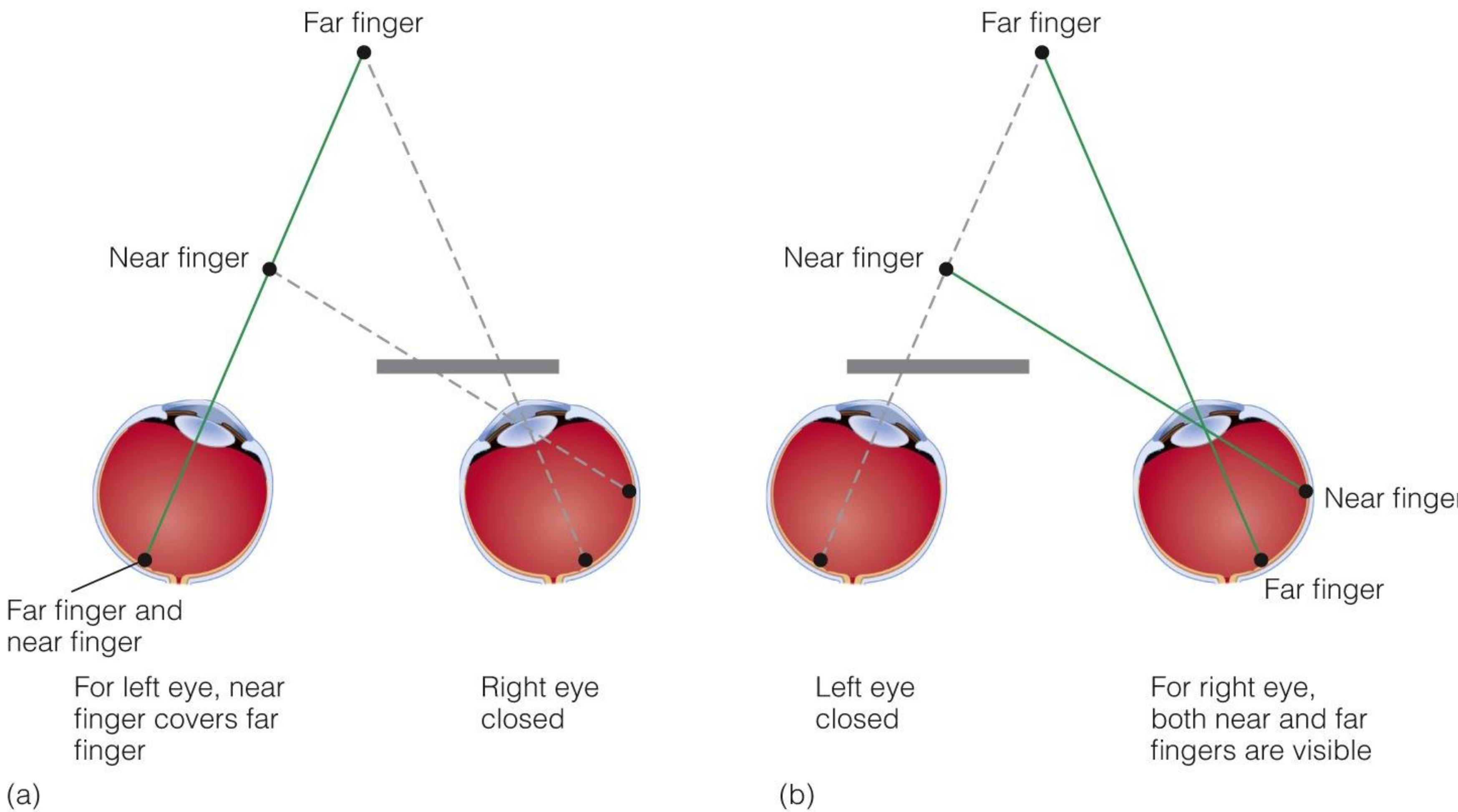
Phil Schermeister/Corbis


Peephole
Twice as far from observer
as the woman on the right.

Depth Perception: Visual Cues

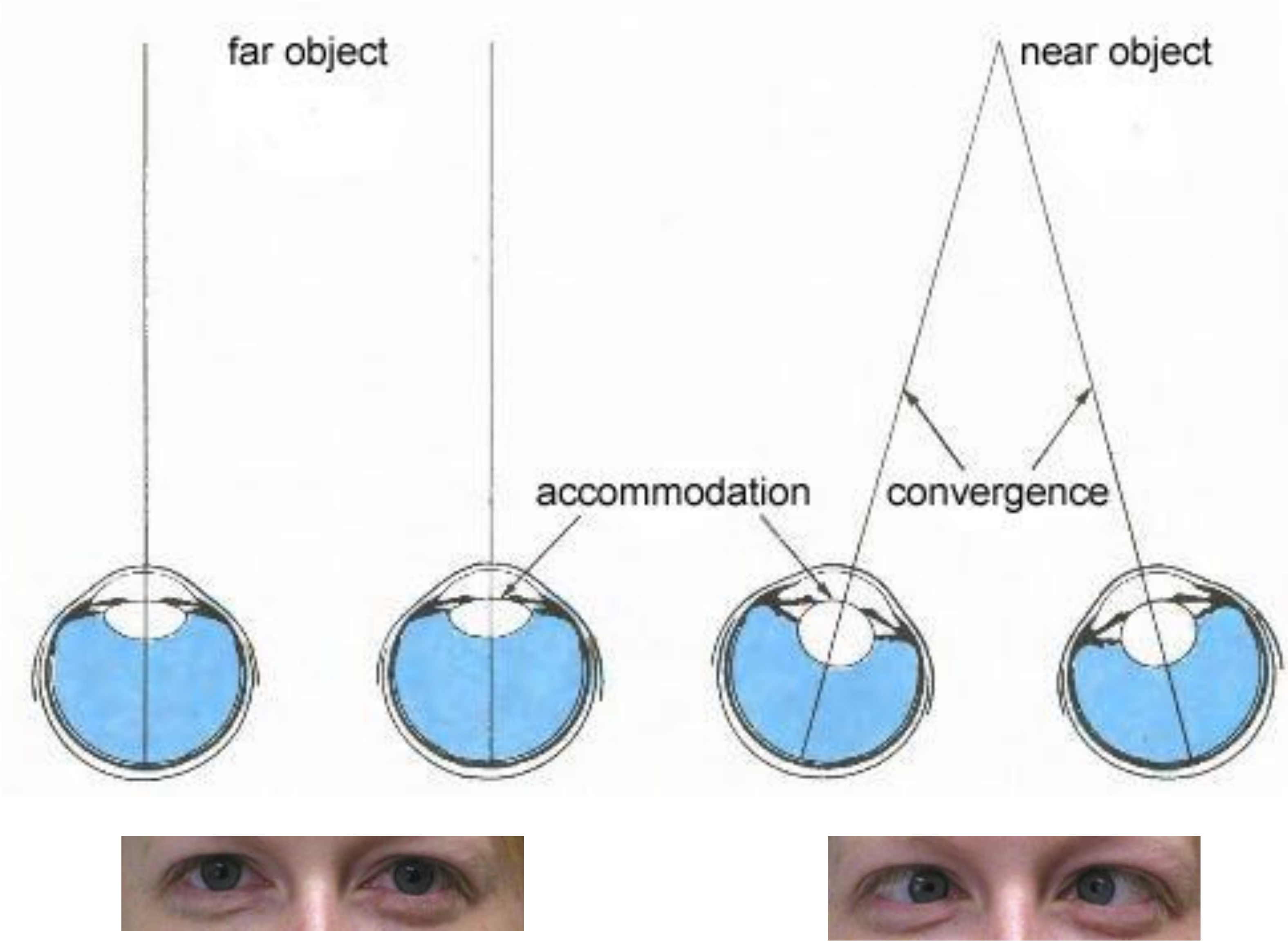
- Binocular disparity
- Oculomotor cues
- Motion cues
- Pictorial cues

Binocular Depth Perception



Depth Perception:
Oculomotor Cues

- Oculomotor cues are based on sensing the position of the eyes and muscle tension
  - Convergence: inward movement of the eyes when we focus on nearby objects
  - Accommodation: change in the shape of the lens when we focus on objects at different distances



Sensation and Perception, E. Bruce Goldstein

Depth Perception:
Motion Cues

- Motion parallax:
  - Close objects in direction of movement glide rapidly past
  - Objects in the distance appear to move slowly
- Deletion and accretion:
  - Objects are covered or uncovered as we move relative to them
  - Covering an object is deletion, uncovering an object is accretion

Depth Perception:
Pictorial Cues (Slide 1 of 3)

- Occlusion:
  - When one object partially covers another behind it
- Relative height:
  - Objects below the horizon that are higher in the field of vision are more distant
  - Objects above the horizon lower in the visual field are more distant

Depth Perception:
Pictorial Cues (Slide 2 of 3)

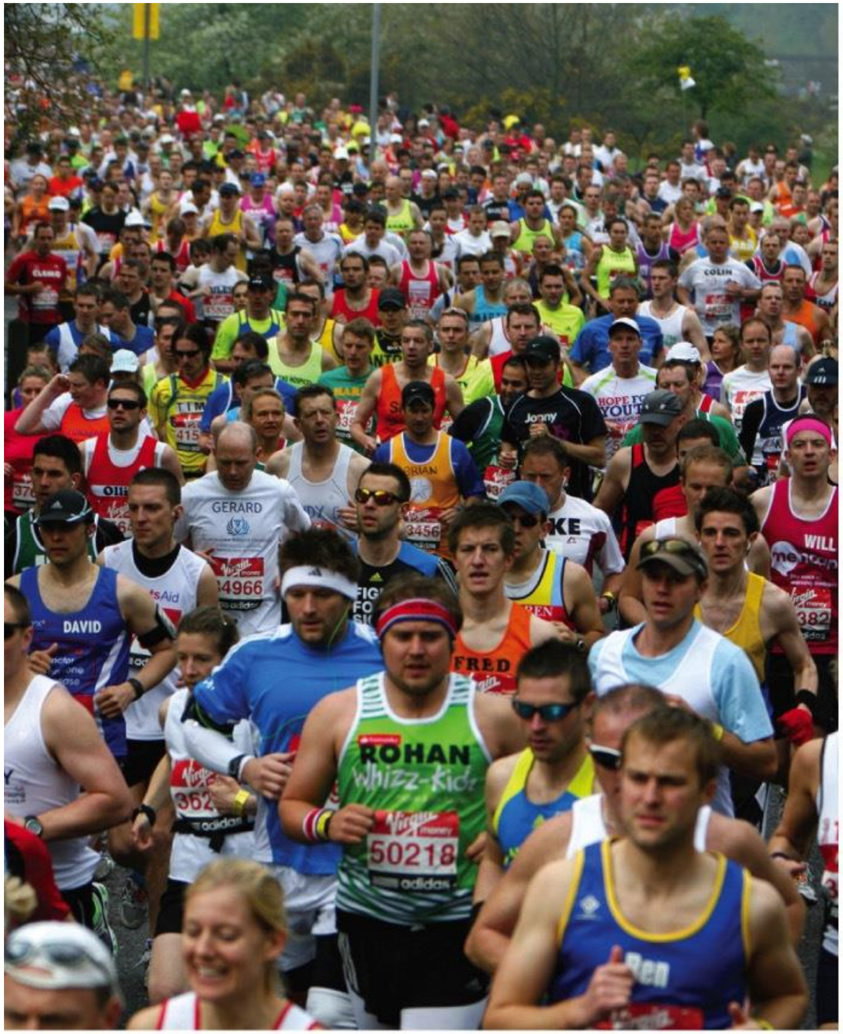
- Relative size
  - When objects are equal size, the closer one will take up more of your visual field
- Perspective convergence
  - Parallel lines appear to come together in the distance
- Familiar size
  - Distance information based on our knowledge of object size

Depth Perception:
Pictorial Cues (Slide 3 of 3)

- Atmospheric perspective
  - Distance objects are fuzzy and have a blue tint
- Texture gradient
  - Equally spaced elements are more closely packed as distance increases
- Shadows
  - Indicate where objects are located
  - Enhance 3D perception of objects

A scene in Tucson, Arizona, containing a number of depth cues:

- **Occlusion** (the cactus on the right occludes the hill, which occludes the mountain)
- **Relative height** (the far motorcycle is higher in the field of view than the closer motorcycle)
- **Relative size** (the far motorcycle and telephone pole are smaller than the near ones)
- **Perspective convergence** (the sides of the road converge in the distance).


DAVID
GERARD
ROHAN
Whizz-Kidz
50218
FRED
WILL
COLIN
34966

Human Visual Perception System is Astonishingly Capable!

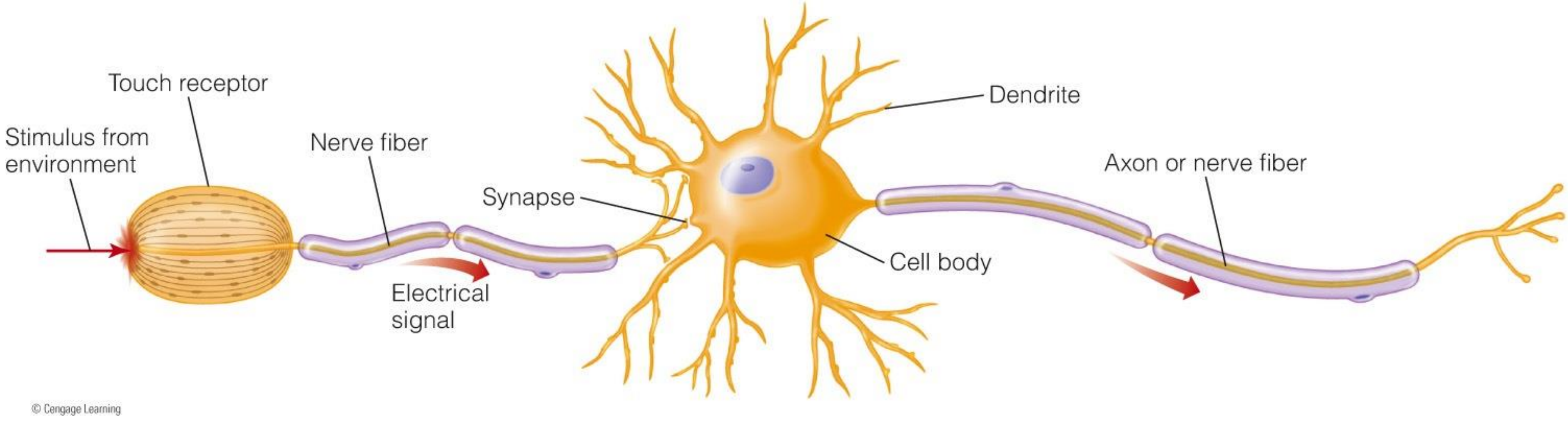
Bev Doolittle

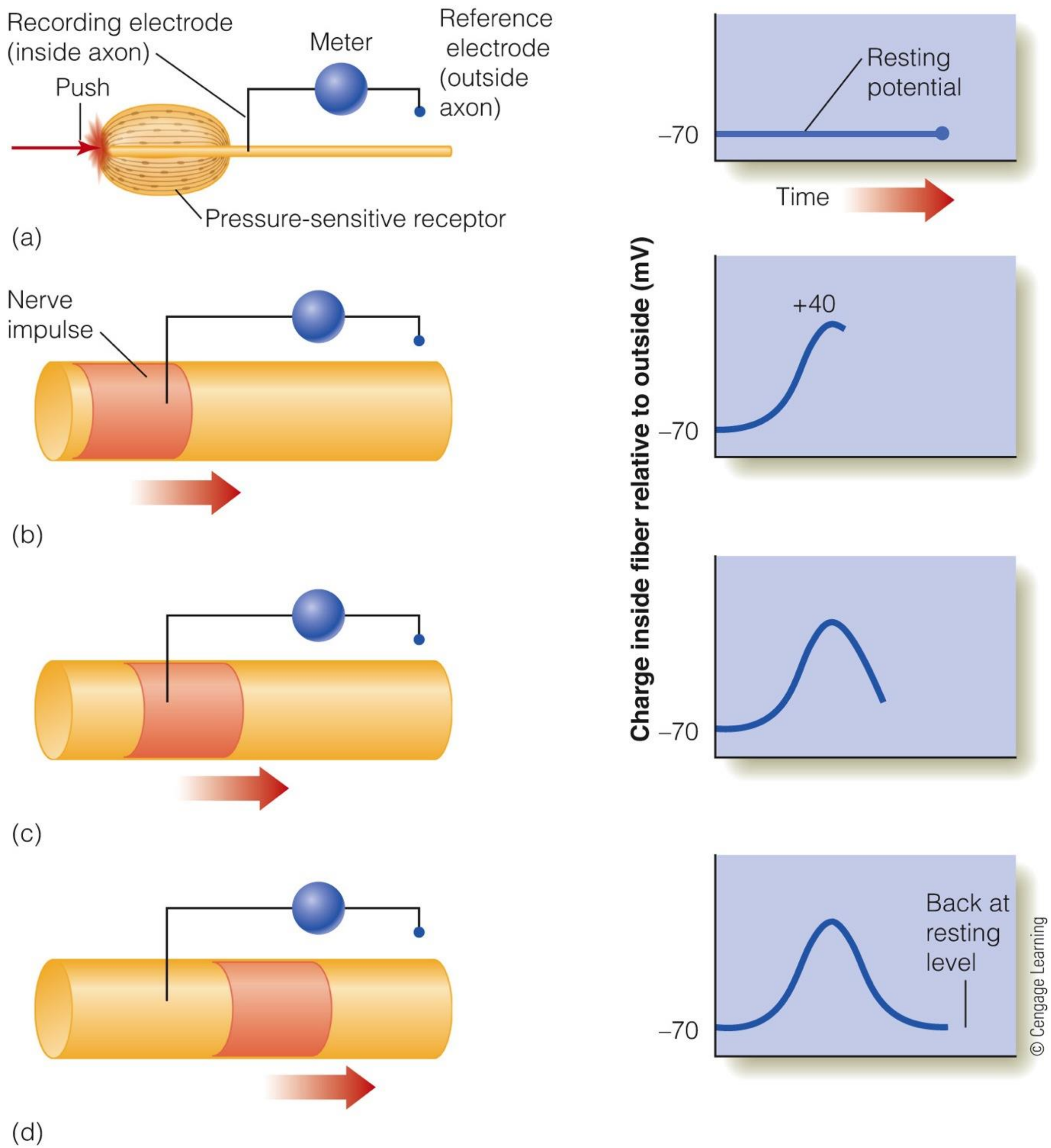
A Lot More Involved Than the Transduction Process

A Lot More Involved Than the Transduction Process

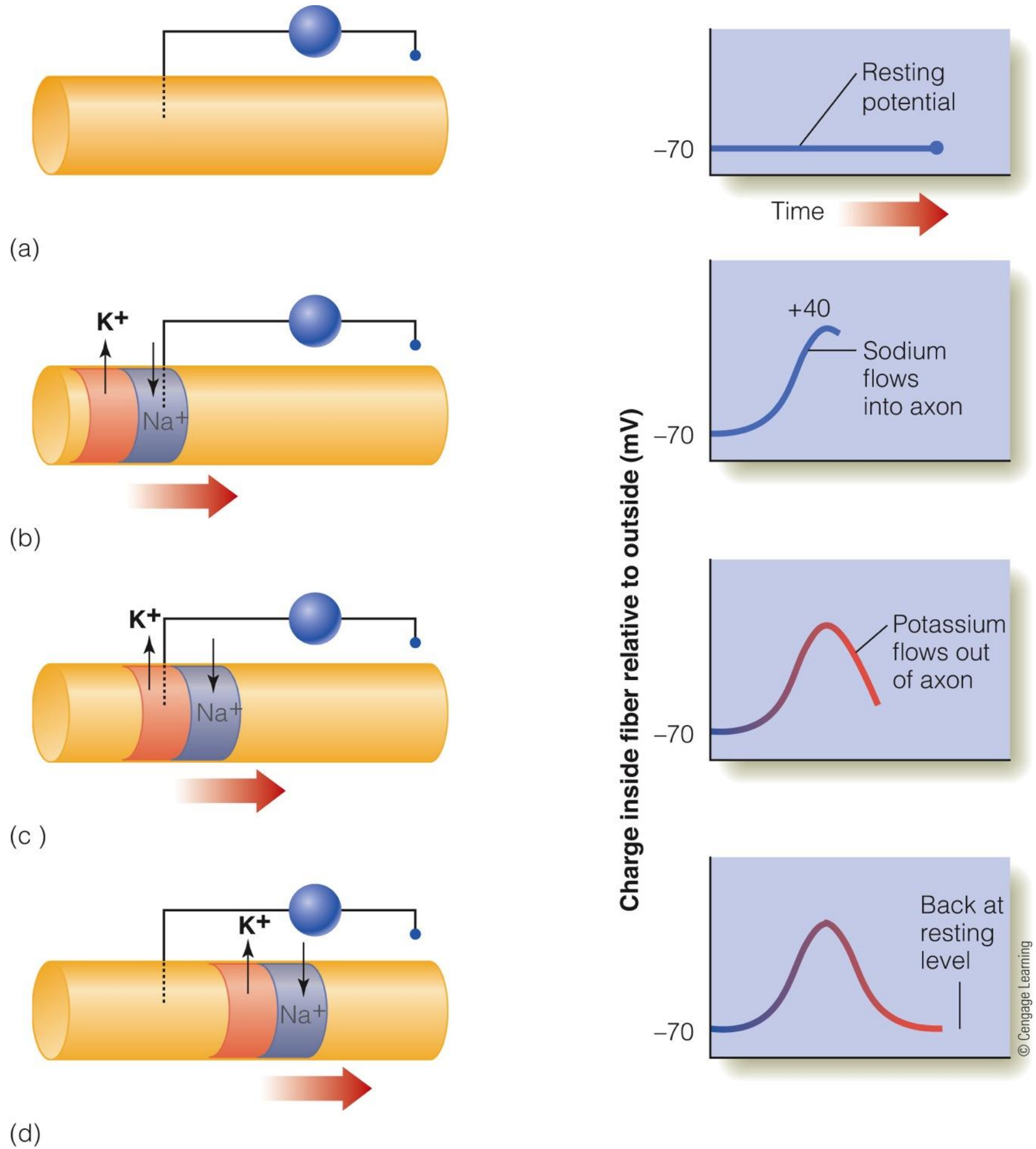
Neurons

- Key components of neurons:
  - Cell body
  - Dendrites
  - Axon or nerve fiber
- Sensory receptors - specialized neurons that respond to specific kinds of energy

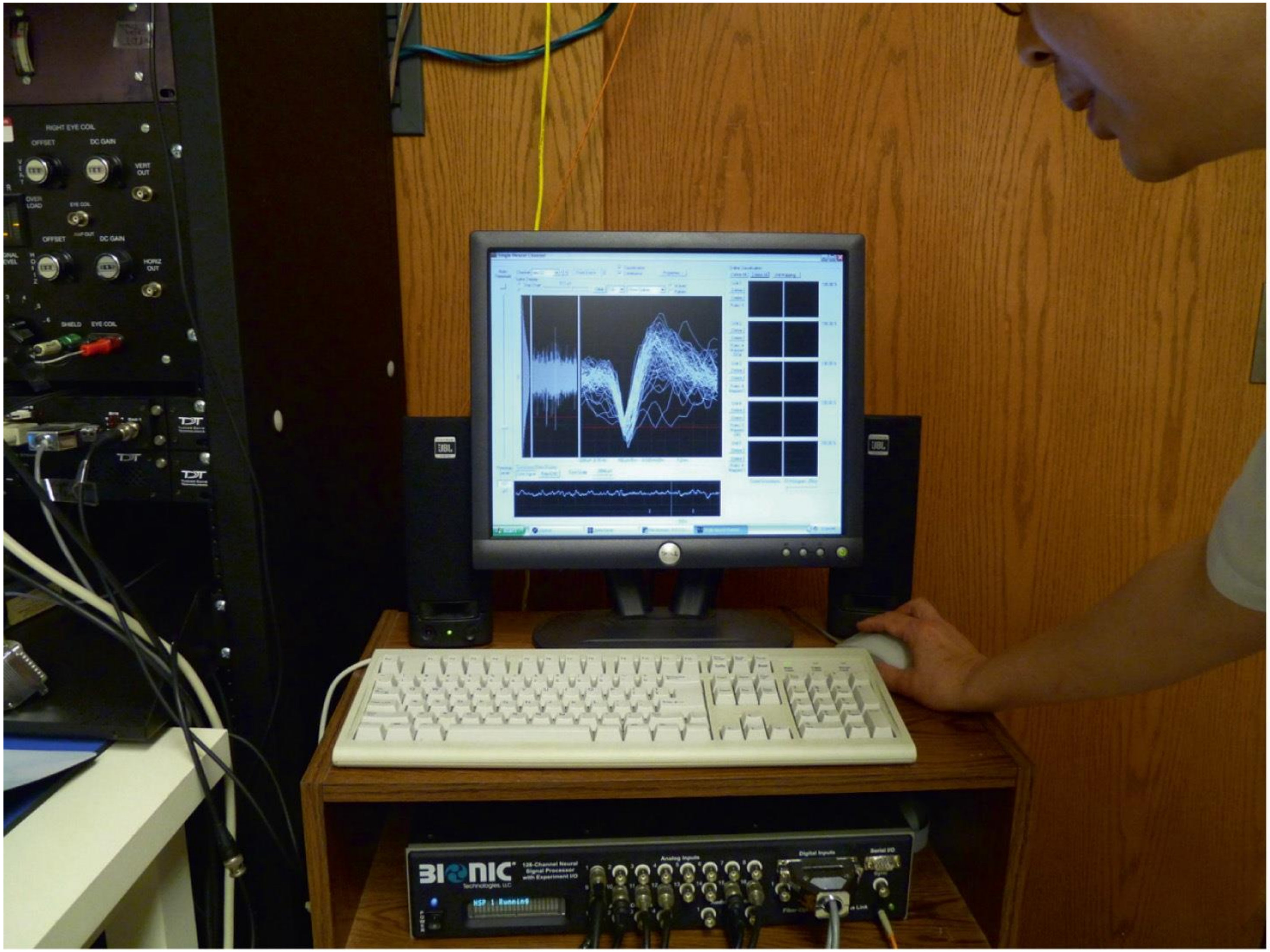




Recording electrode (inside axon)
Meter
Reference electrode (outside axon)
Push
Pressure-sensitive receptor
(a)
Resting potential
–70
Time
Nerve impulse
(b)
+40
–70
(c)
–70
(d)
Back at resting level
–70
Charge inside fiber relative to outside (mV)
© Cengage Learning



(a)
(b)
(c)
(d)
K+
Na+
Resting
potential
–70
Time
+40
Sodium
flows
into axon
Potassium
flows out
of axon
Back at
resting
level
Charge inside fiber relative to outside (mV)
© Cengage Learning

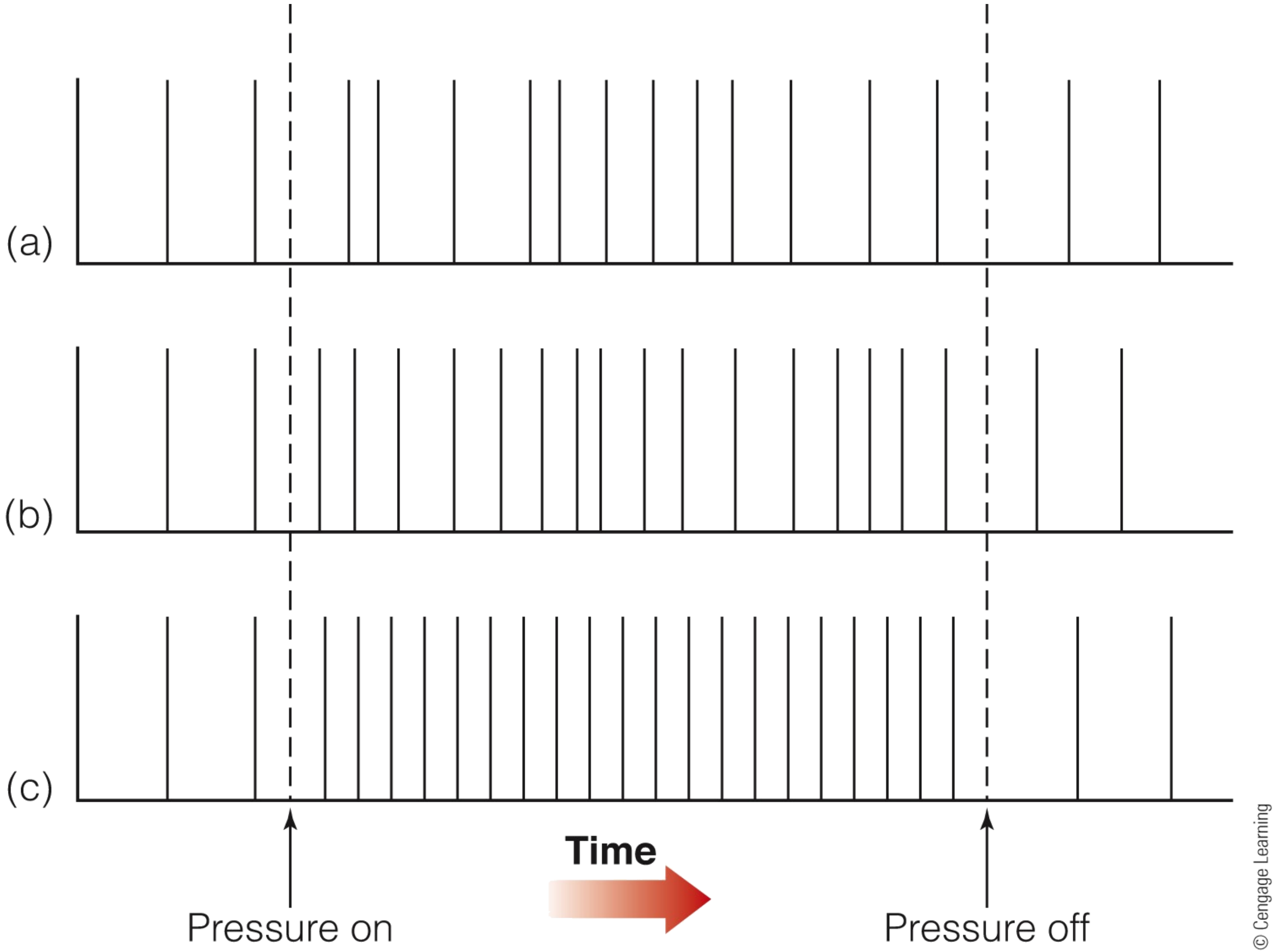


Bruce Goldstein

Photographed in Tai Sing Lee’s laboratory at Carnegie Mellon University

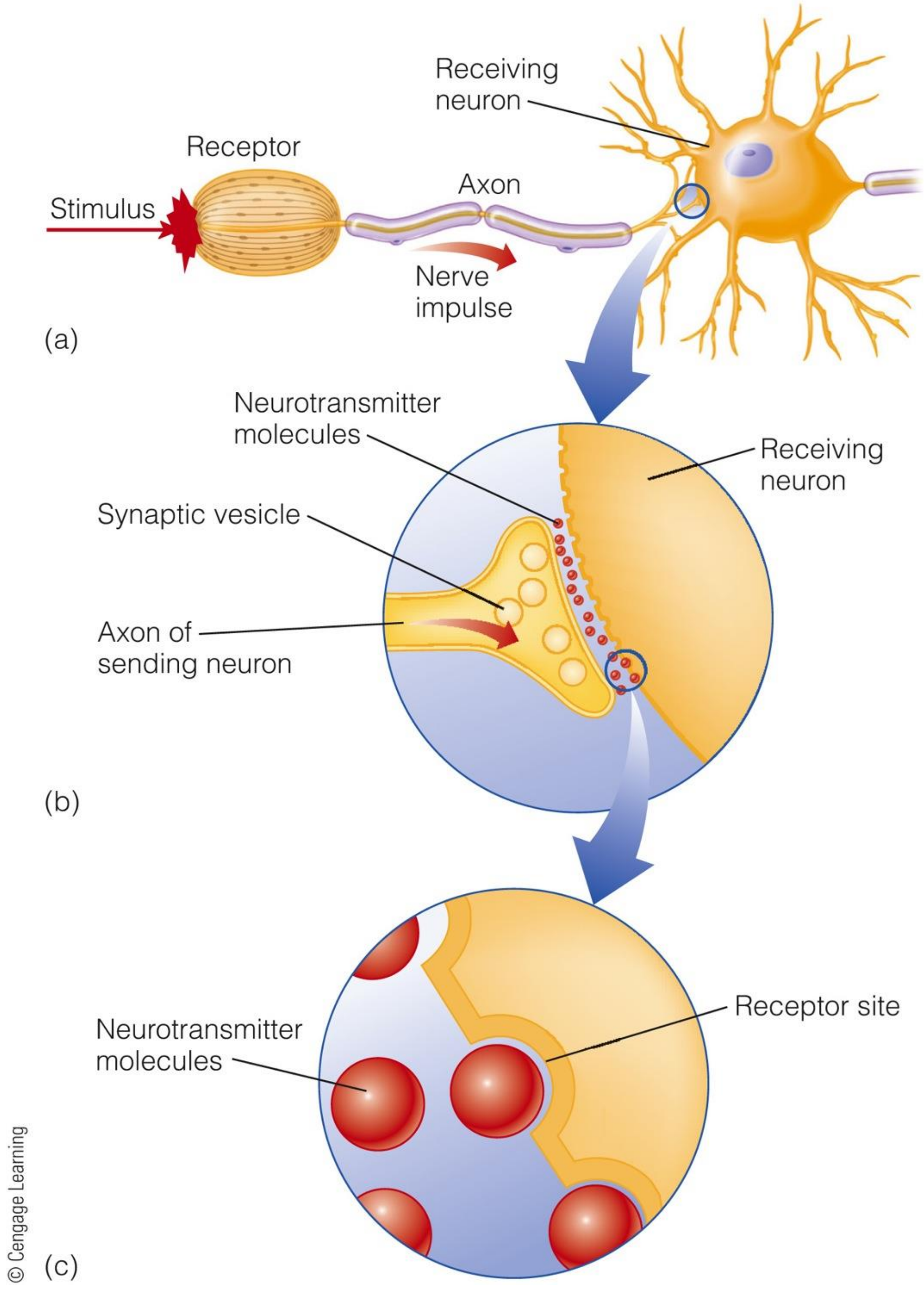
Basic Properties of Action Potentials

- Action potentials:
  - show propagated response.
  - remain the same size regardless of stimulus intensity.
  - increase in rate to increase in stimulus intensity.
  - have a refractory period of 1 ms - upper firing rate is 500 to 800 impulses per second.
  - show spontaneous activity that occurs without stimulation.



Response of a nerve fiber to (a) soft, (b) medium, and (c) strong stimulation.

Transmitting Information Across the Gap



Transmitting Information Across the Gap - continued

- Excitatory transmitters
  - Neuron becomes more positive
  - Increases the likelihood of an action potential
- Inhibitory transmitters
  - Neuron becomes more negative
  - Decreases the likelihood of an action potential

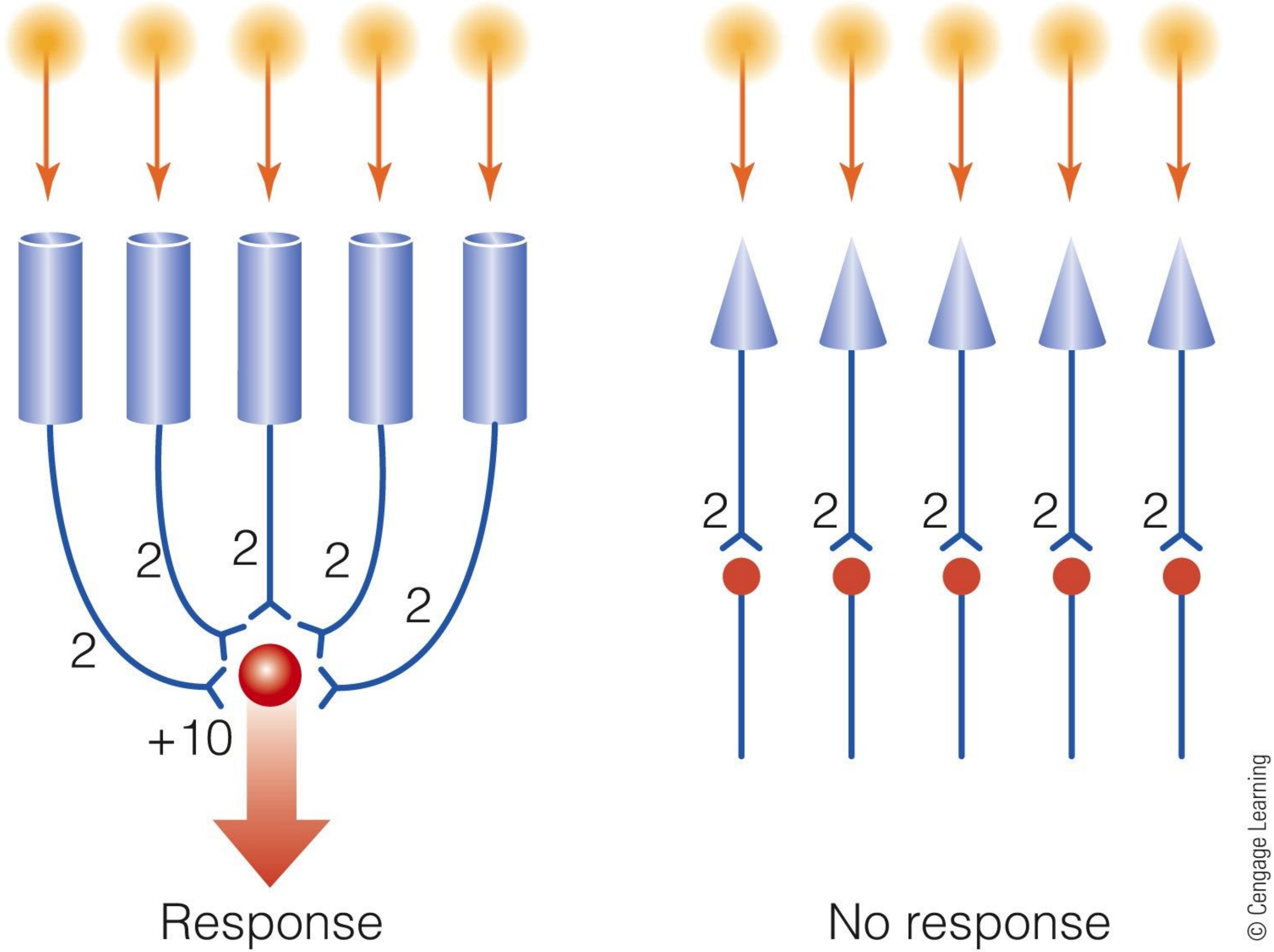
Back to the Retina: Convergence

- 126 million rods and cones converge to 1 million ganglion cells.

- Higher convergence of rods than cones
  - Average of 120 rods to one ganglion cell
  - Average of 6 cones to one ganglion cell
  - Cones in fovea have one to one relation to ganglion cells

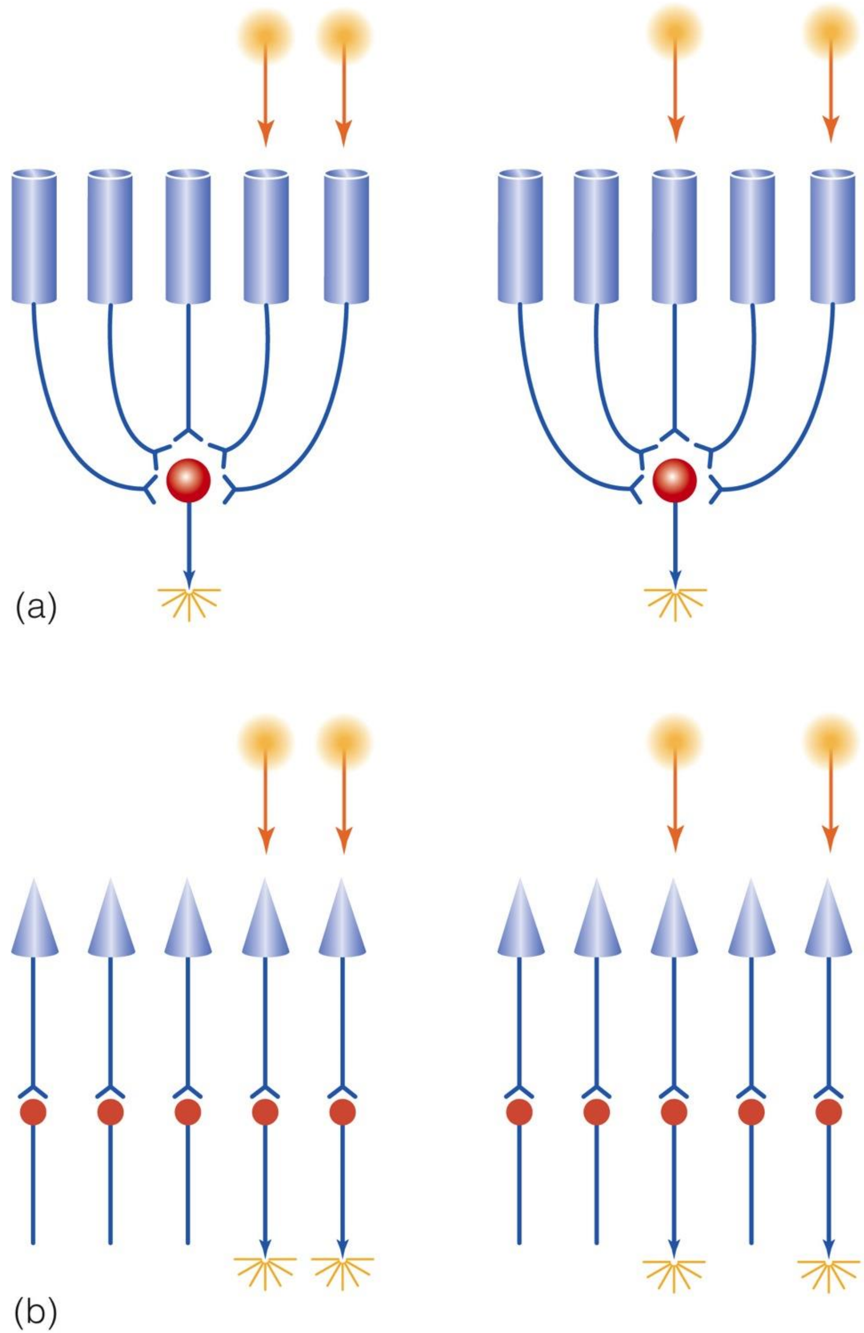
Convergence Causes the Rods to Be More Sensitive Than the Cones

- Rods are more sensitive to light than cones.
  - Rods take less light to respond
  - Rods have greater convergence which results in summation of the inputs of many rods into ganglion cells increasing the likelihood of response.
  - Trade-off is that rods cannot distinguish detail


2
2
2
2
2
+10
Response
2
2
2
2
2
No response
© Cengage Learning

Lack of Convergence Causes the Cones to Have Better Acuity That the Rods

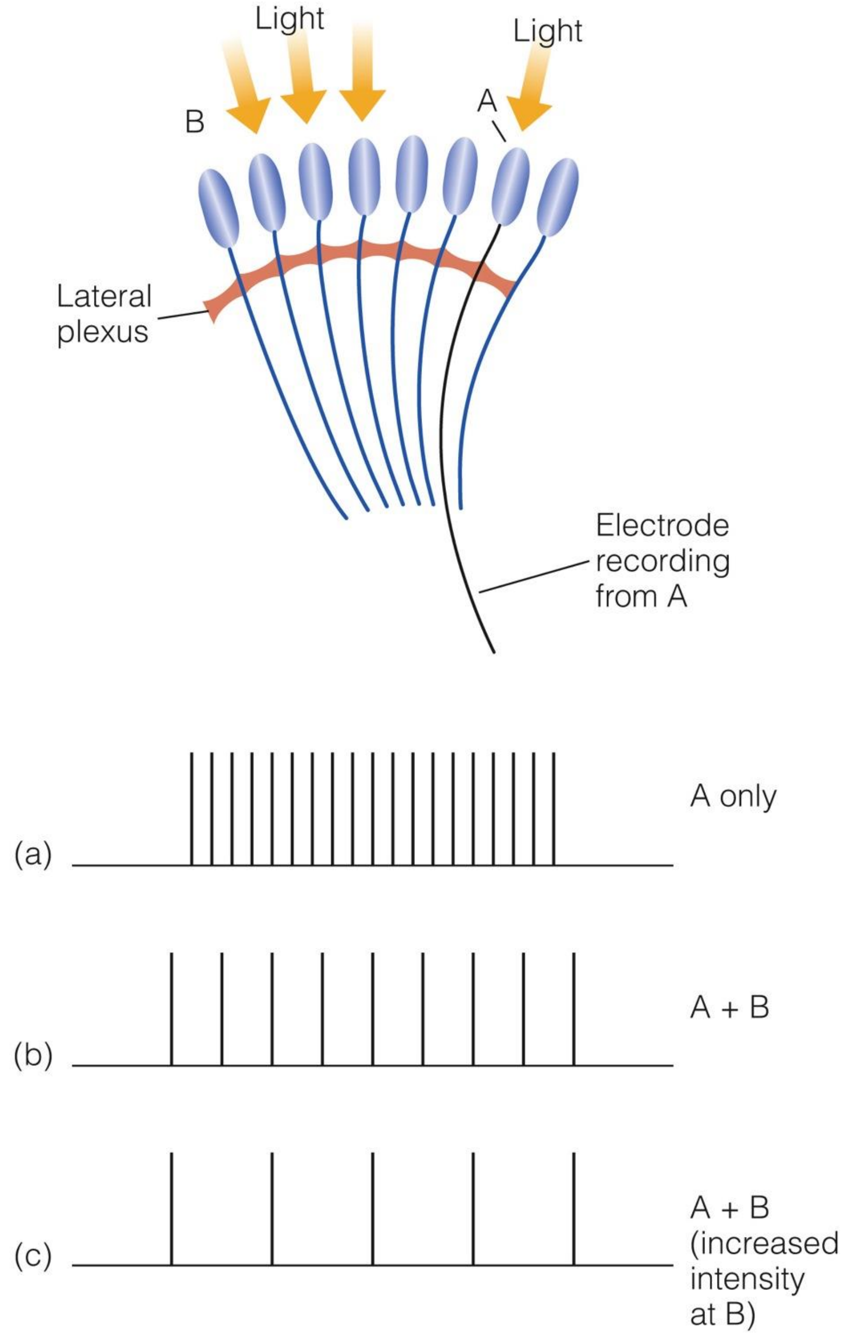
- All-cone foveal vision results in high visual acuity
  - One-to-one wiring leads to ability to discriminate details
  - Trade-off is that cones need more light to respond than rods


(a)
(b)

Bruce Goldstein

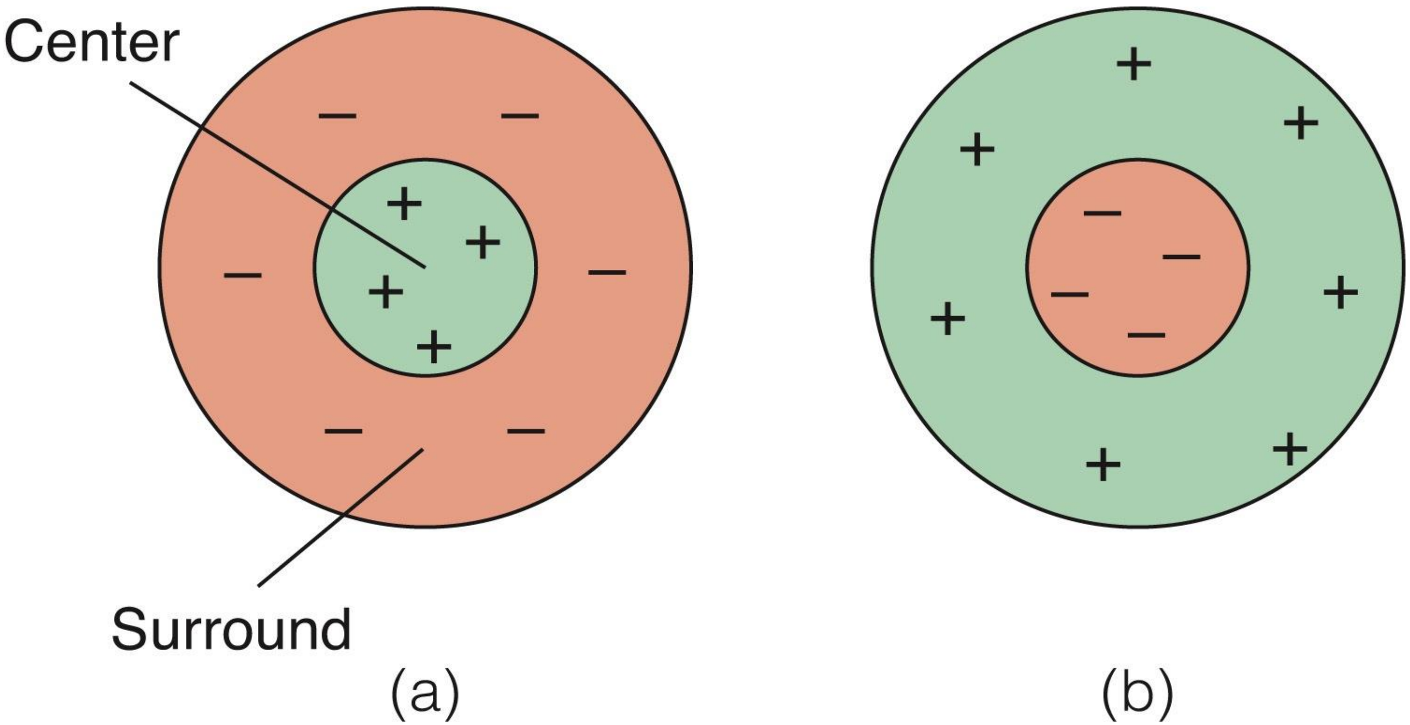
Neural Processing and Perception

Lateral Inhibition



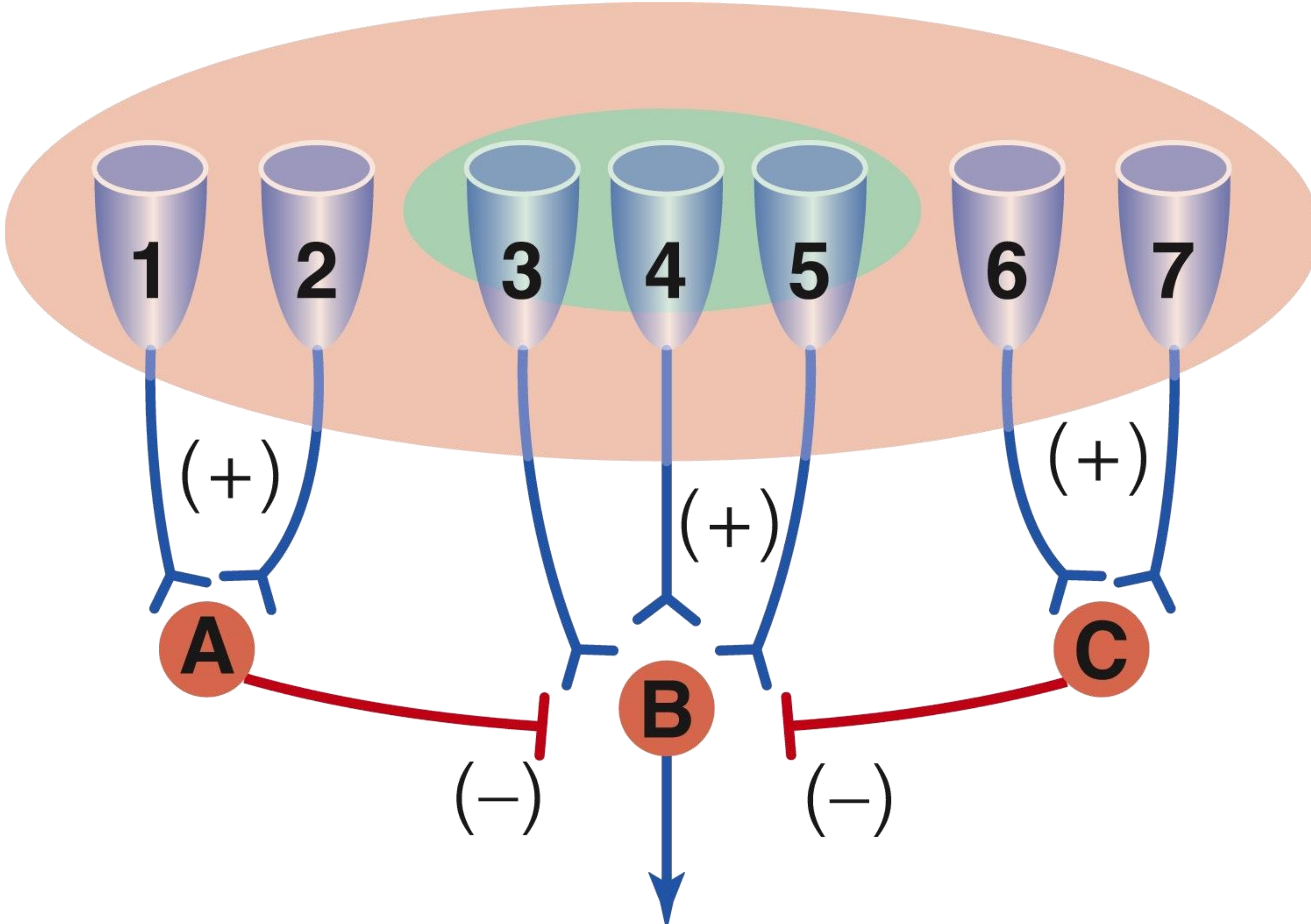
Hartline
Wagner
Ratliff
(1956)

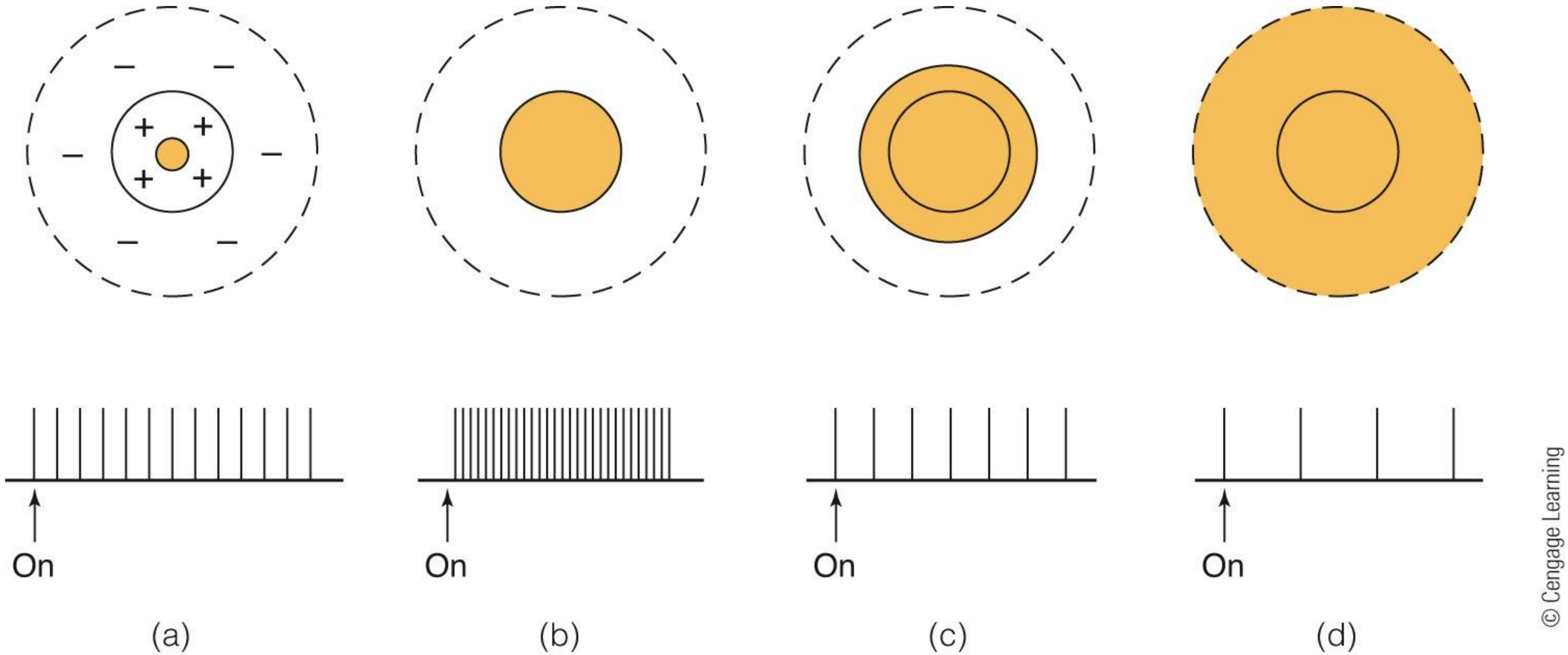
Center-Surround Receptive Fields



(a) Excitatory center, inhibitory surround;
(b) Inhibitory center, excitatory surround.

Kuffler, ‘53
Barlow, ‘57
Hubel & Weisel, ‘65

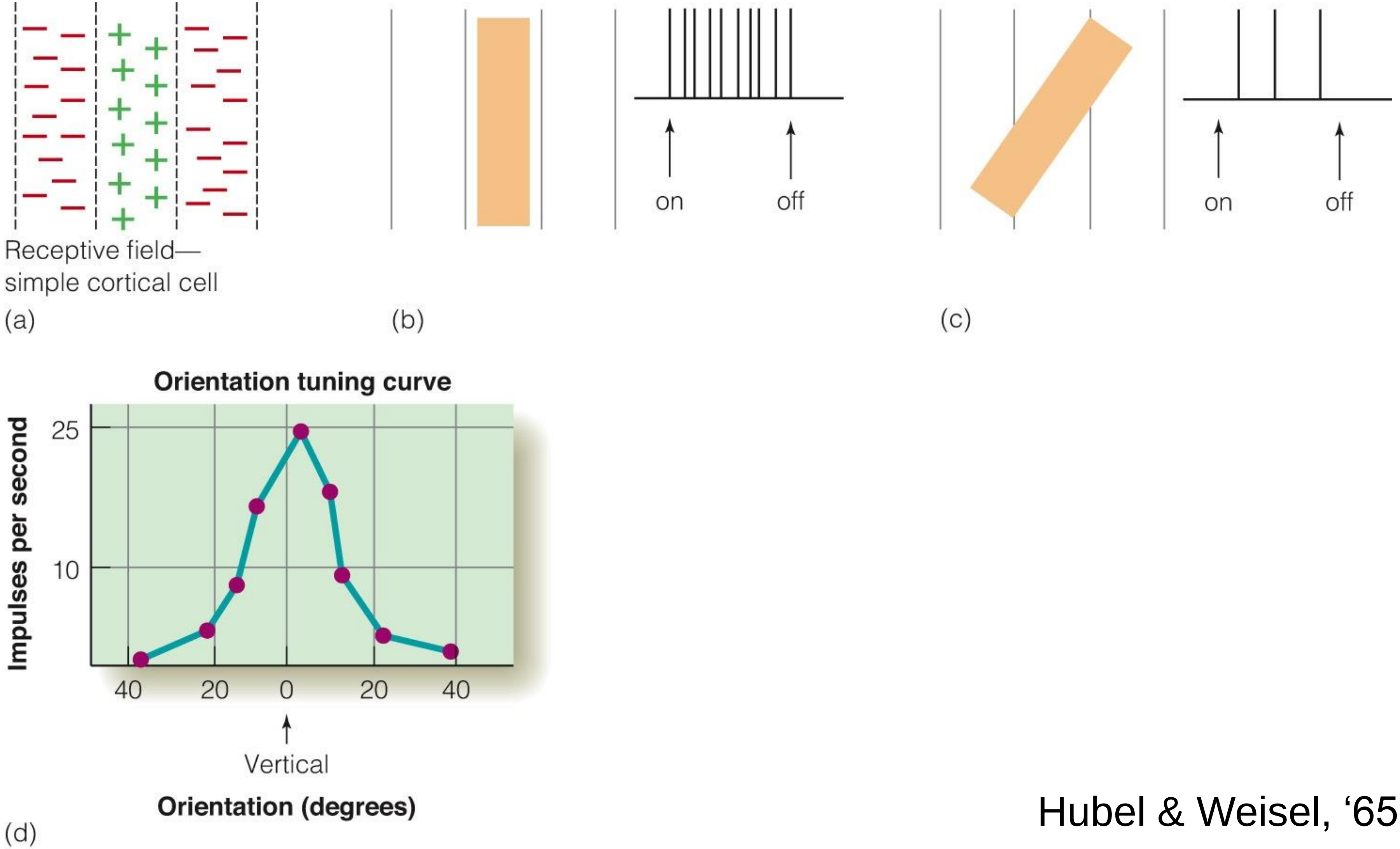

1
2
3
4
5
6
7
(+)
(+)
(+)
A
B
C
(–)
(–)



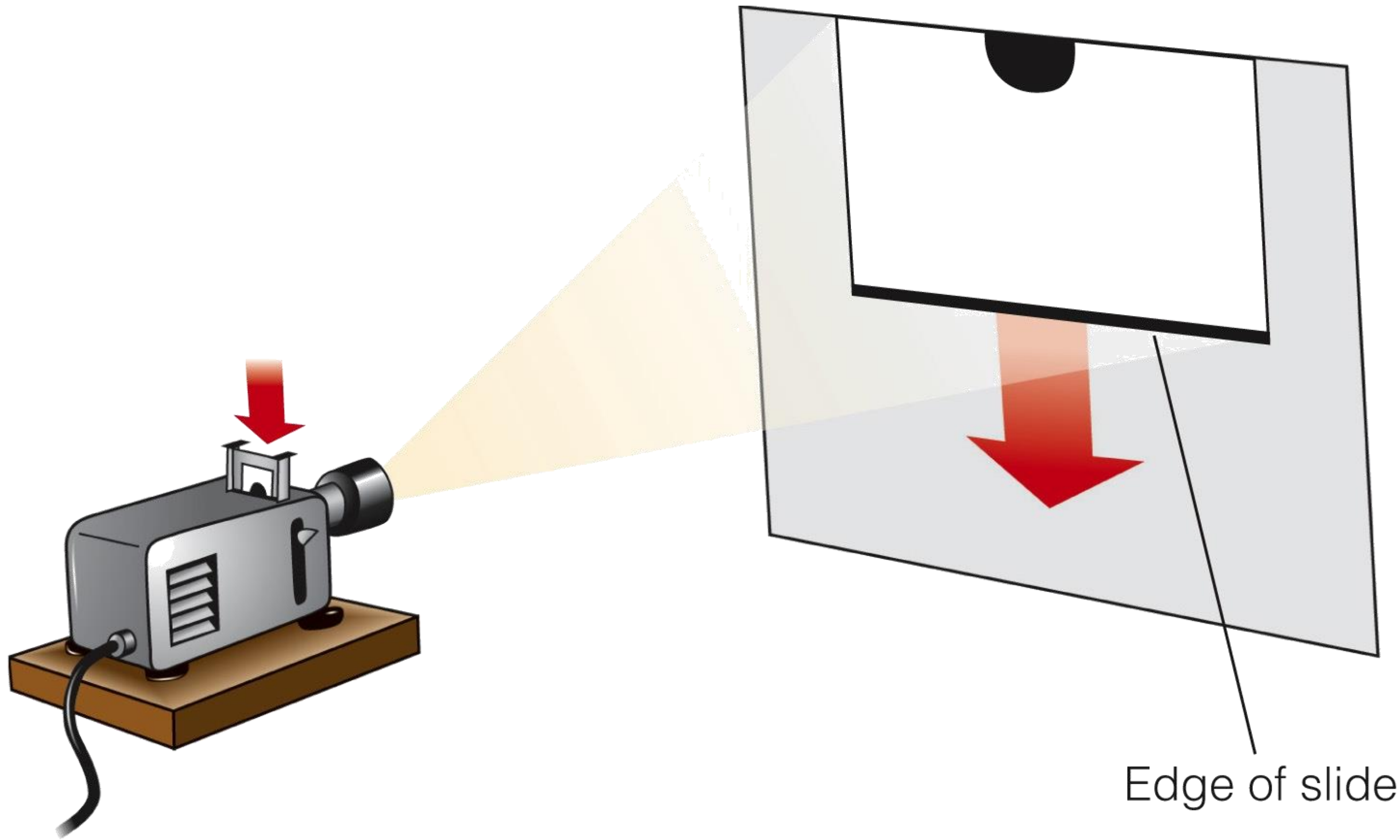
Response of an excitatory-center, inhibitory-surround receptive field as stimulus size is increased.

Shading indicates the area stimulated with light.

The receptive field of a simple cortical cell

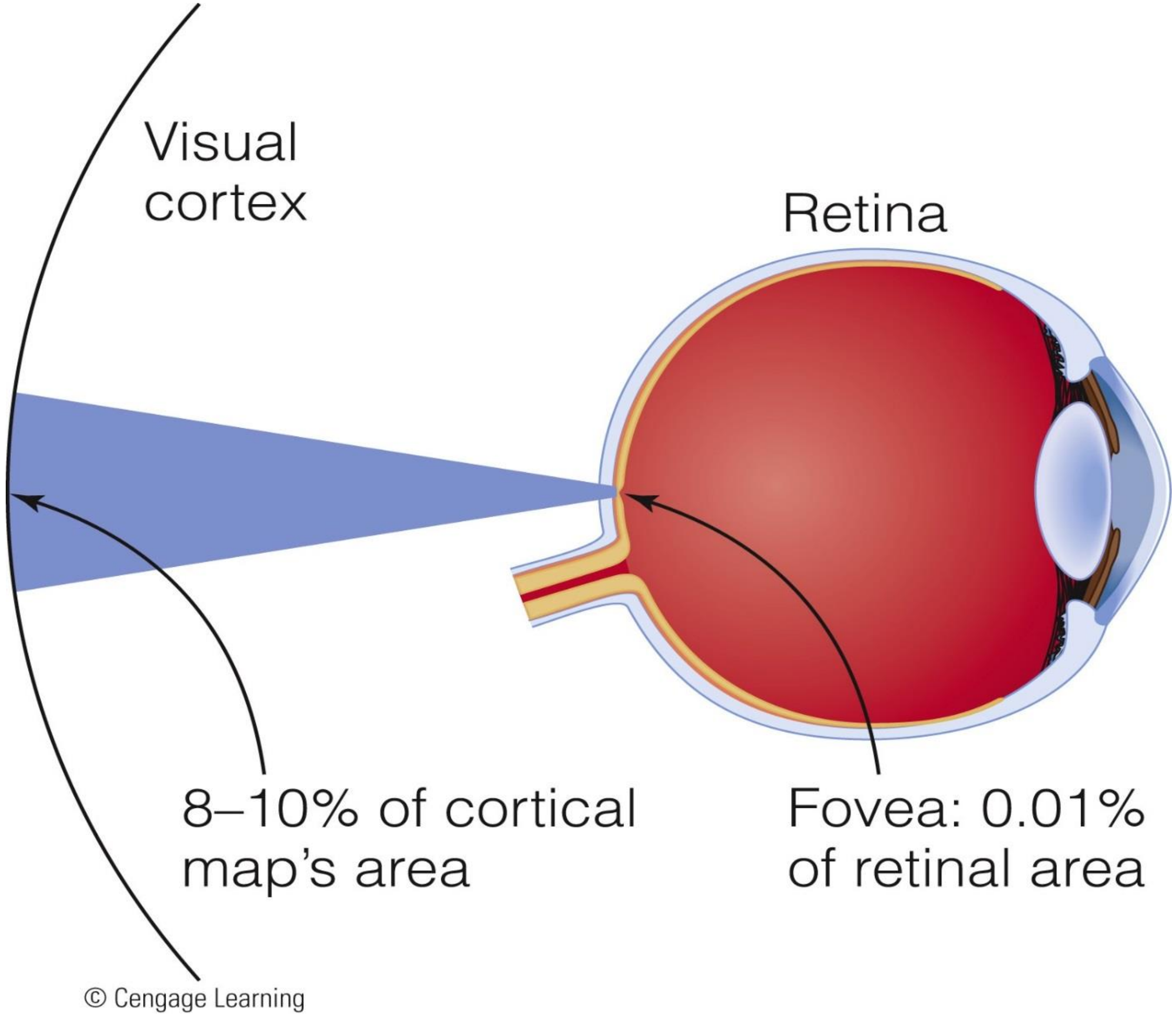


Hubel & Weisel, ‘65

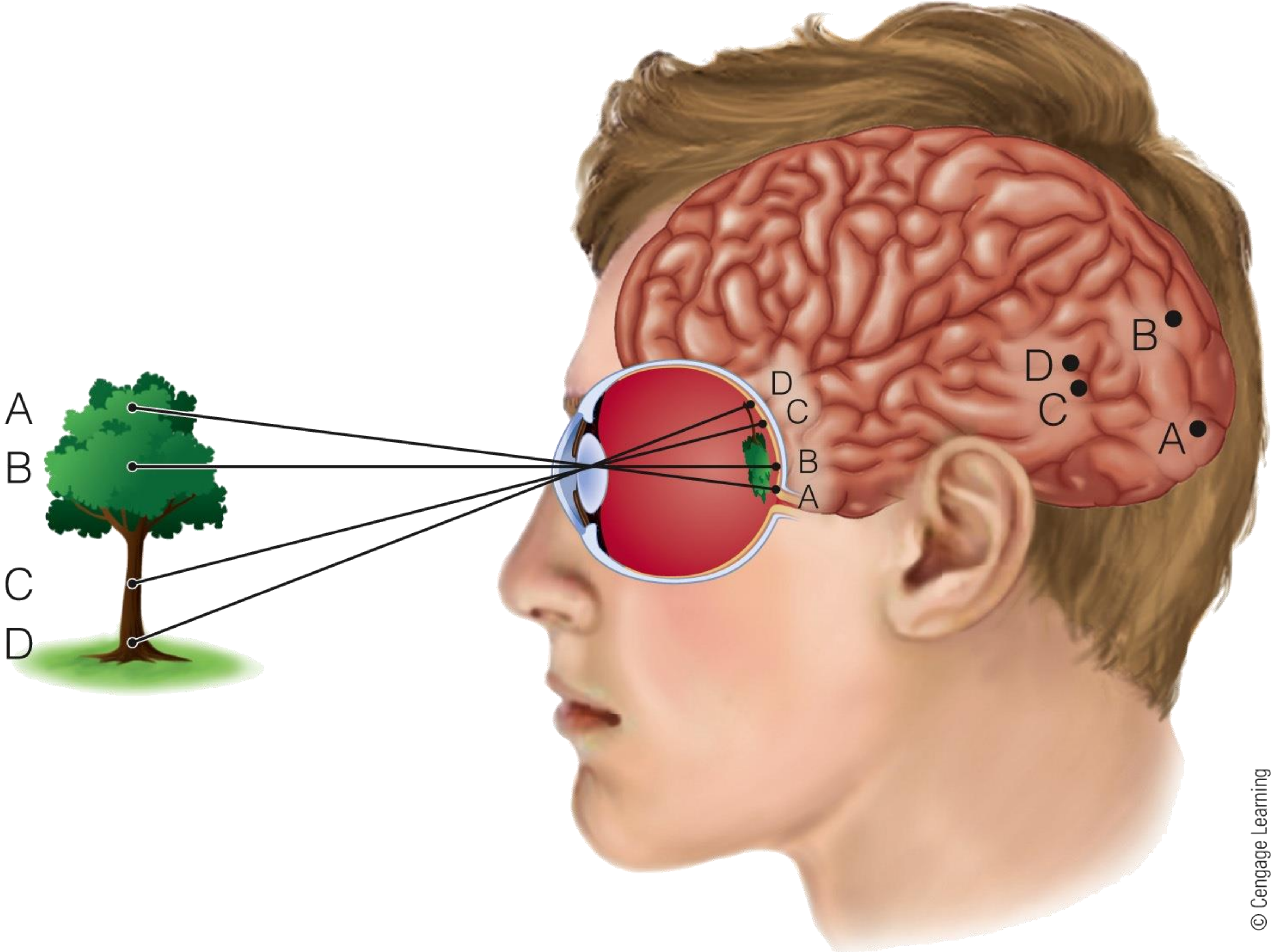


When Hubel and Wiesel dropped a slide into their slide projector, the image of the edge of the slide moving down unexpectedly triggered activity in a cortical neuron of a cat. (Reports: 1962, Nobel Prize: 1981)

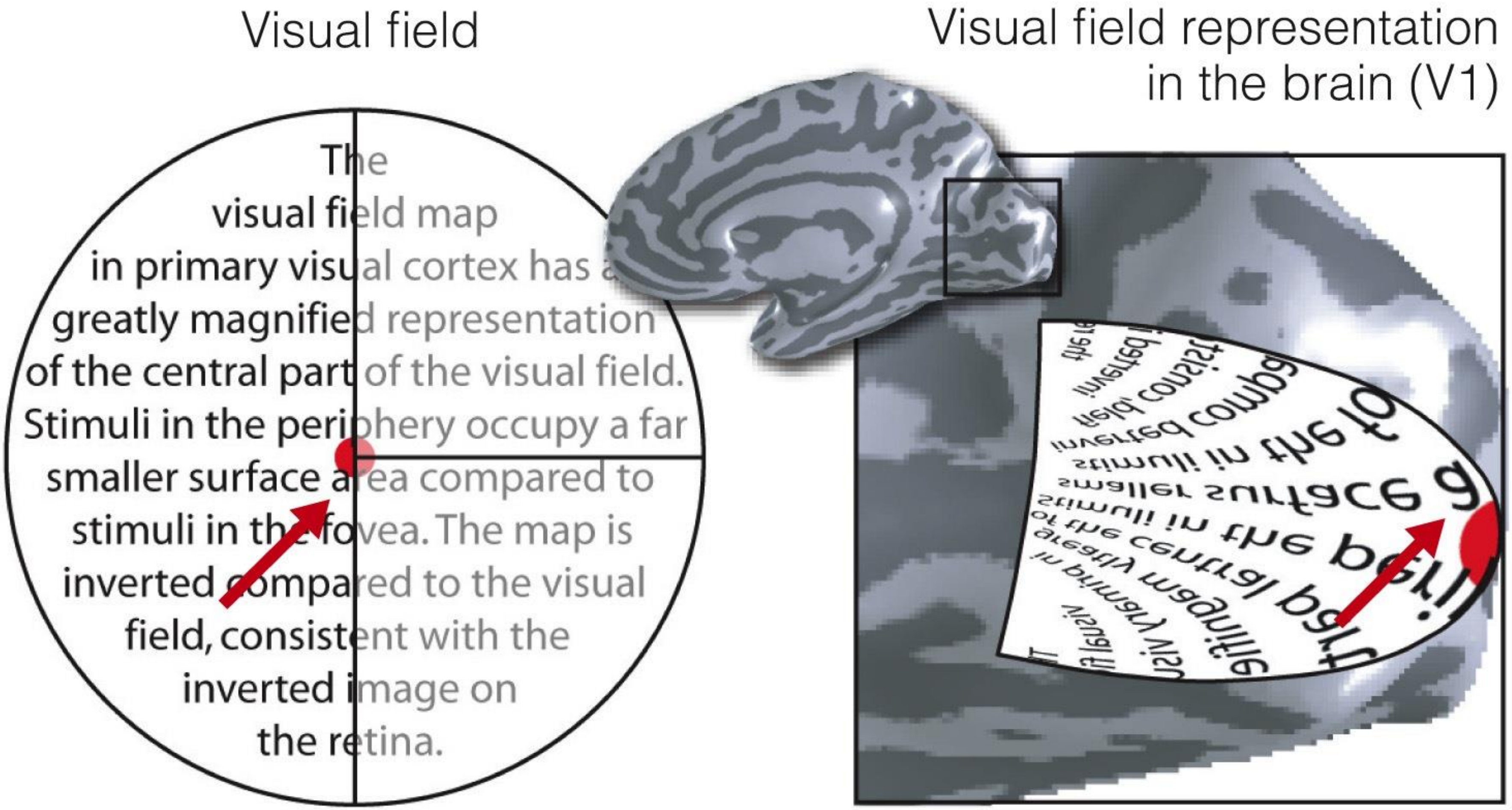
Cortical Organization



The magnification factor in the visual system. The small area of the fovea is represented by a large area on the visual cortex.


A
B
C
D
D
C
B
A
B
D
C
A
© Cengage Learning

Cortical Magnification

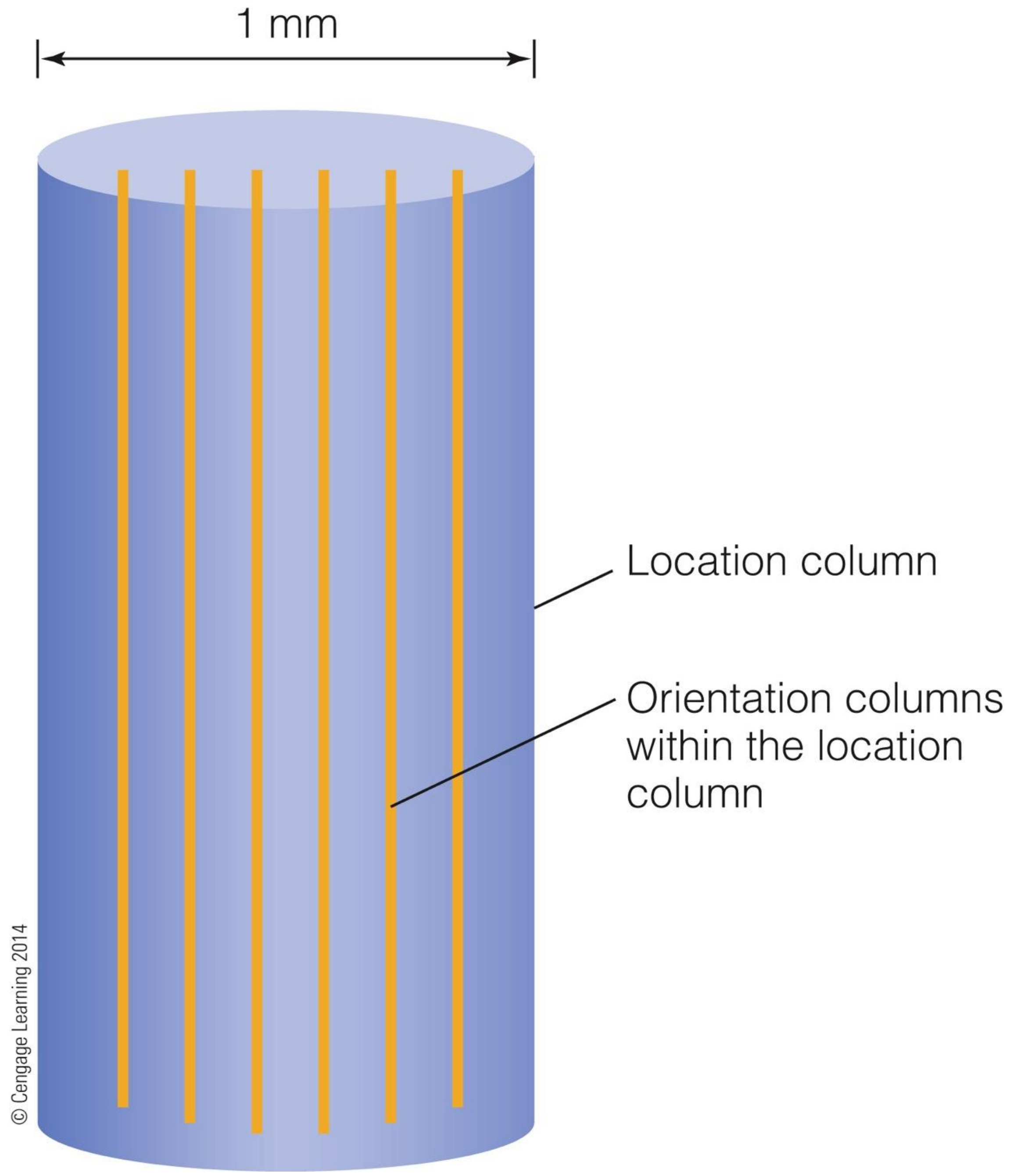


Reprinted from Wandell, B. A., Dumoulin, S. O., & Brewer, A. A. (2009). Visual areas in humans. In L. Squire (Ed.), *Encyclopedia of neuroscience*, Fig. 6, with permission from Elsevier.

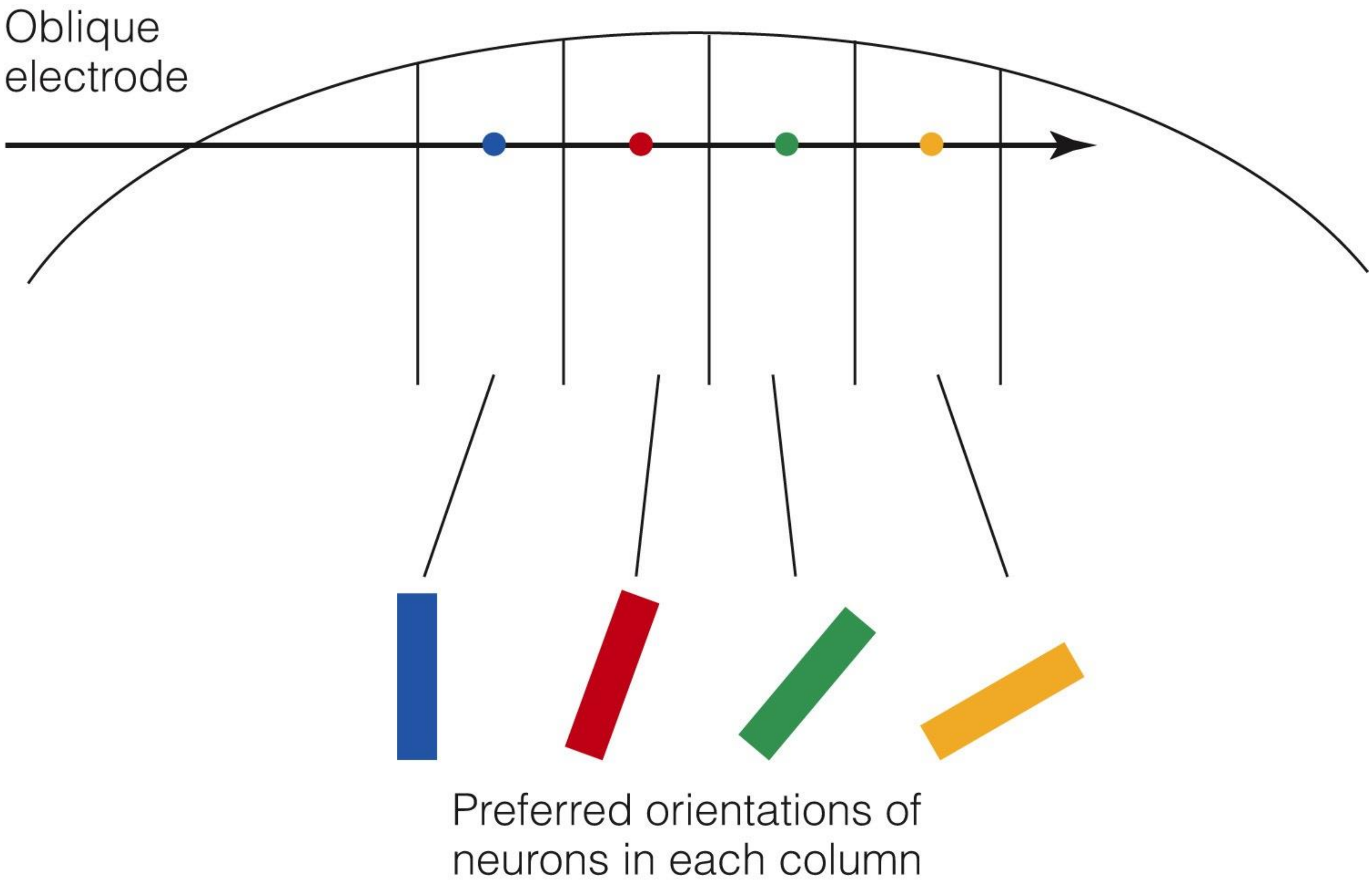
Wandell, 2009

The Cortex is Organized in Columns

- Location columns
  - Receptive fields at the same location on the retina are within a column
- Orientation columns
  - Neurons within columns fire maximally to the same orientation of stimuli
  - Adjacent columns change preference in an orderly fashion
  - 1 millimeter across the cortex represents entire range of orientation

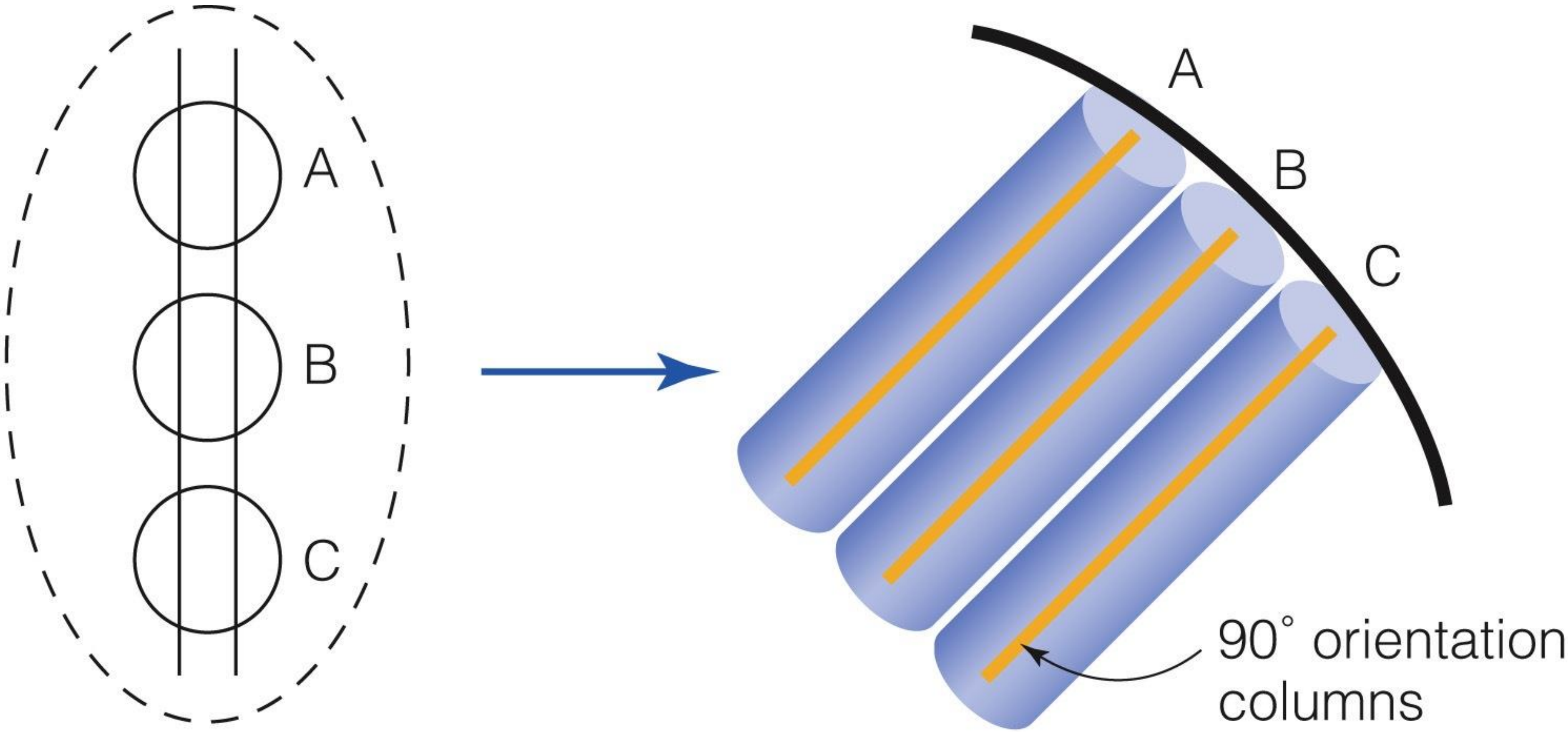

1 mm
Location column
Orientation columns within the location column

Orientation Columns

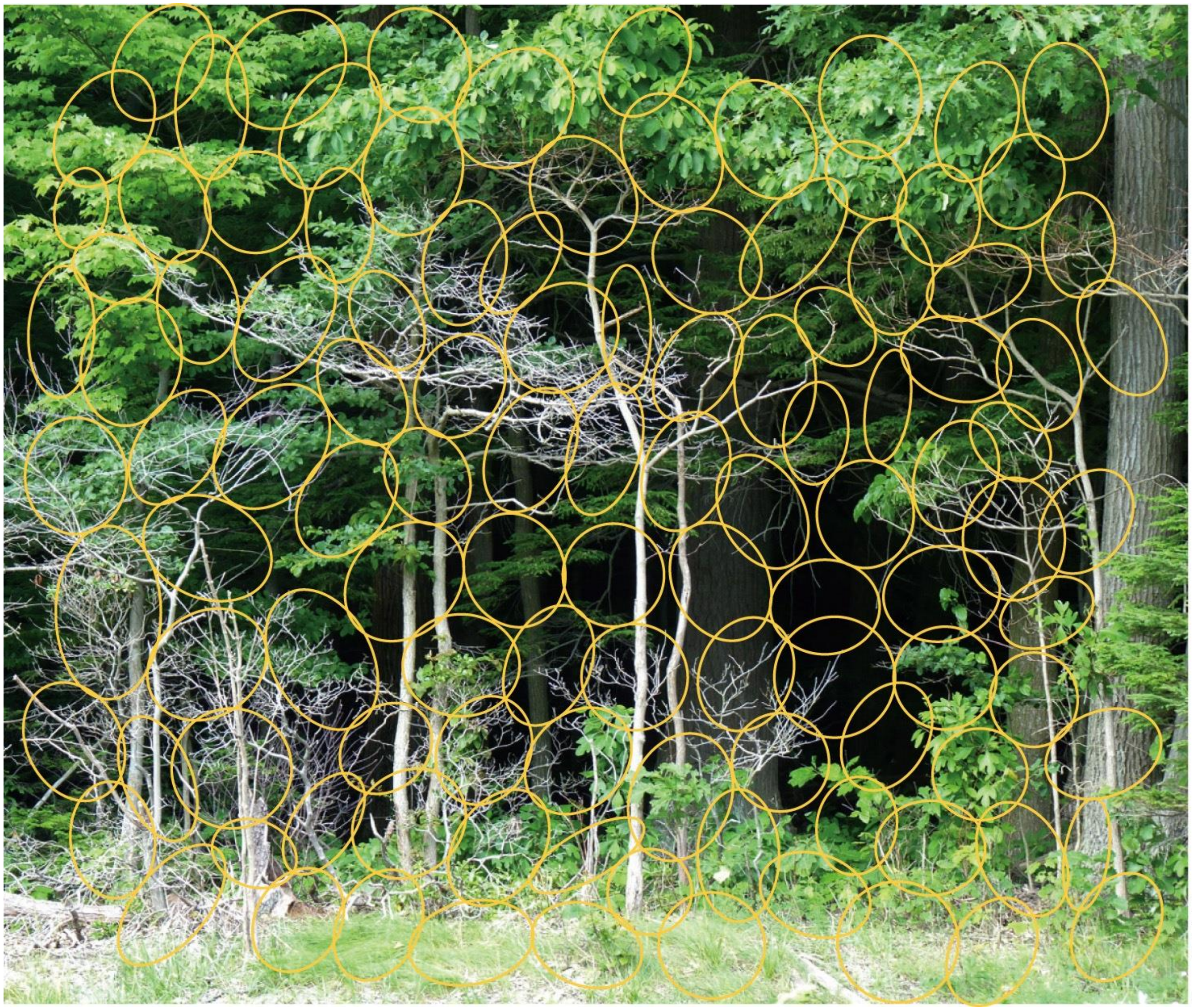


Bruce Goldstein

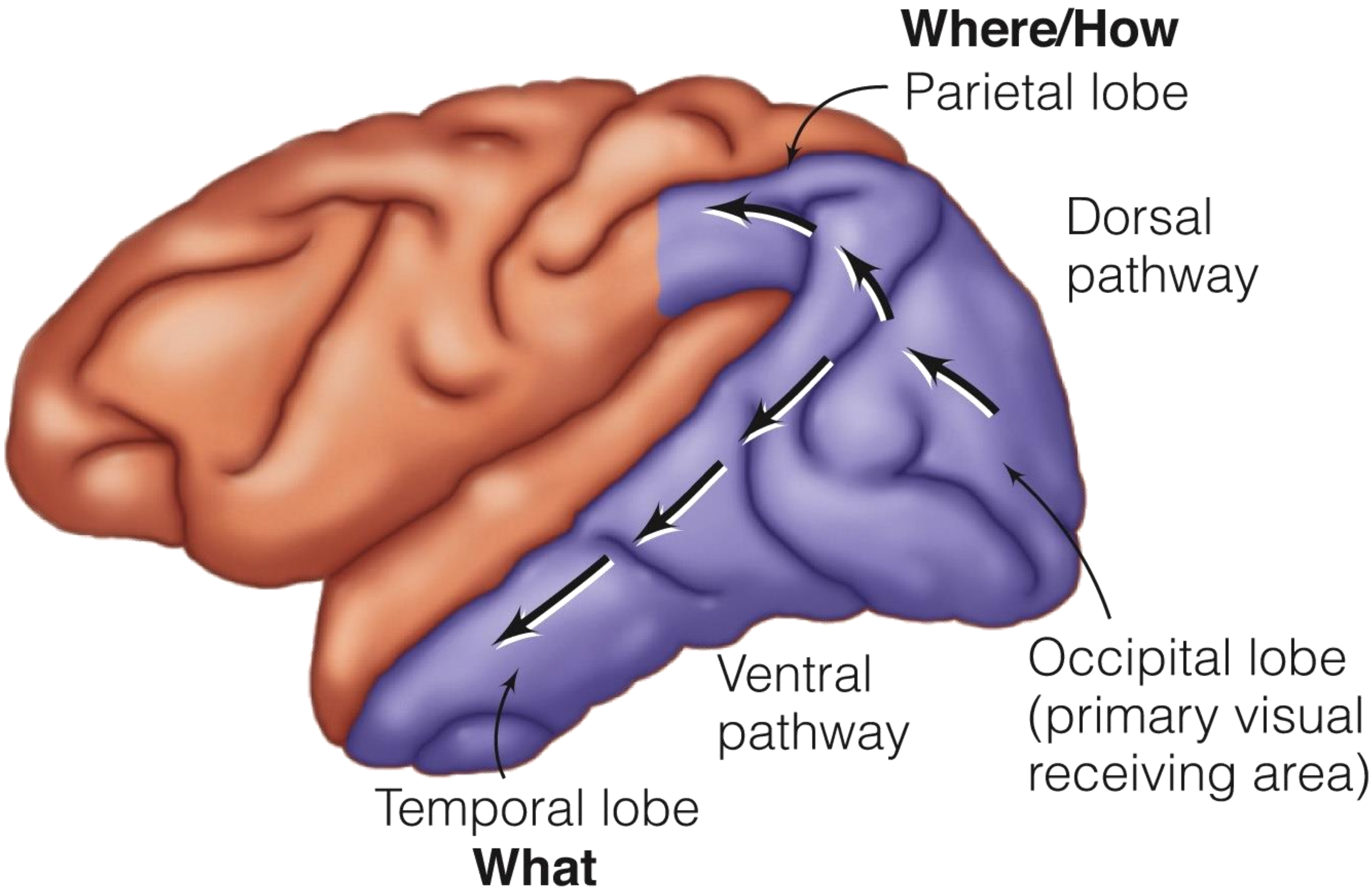
(a) (b)

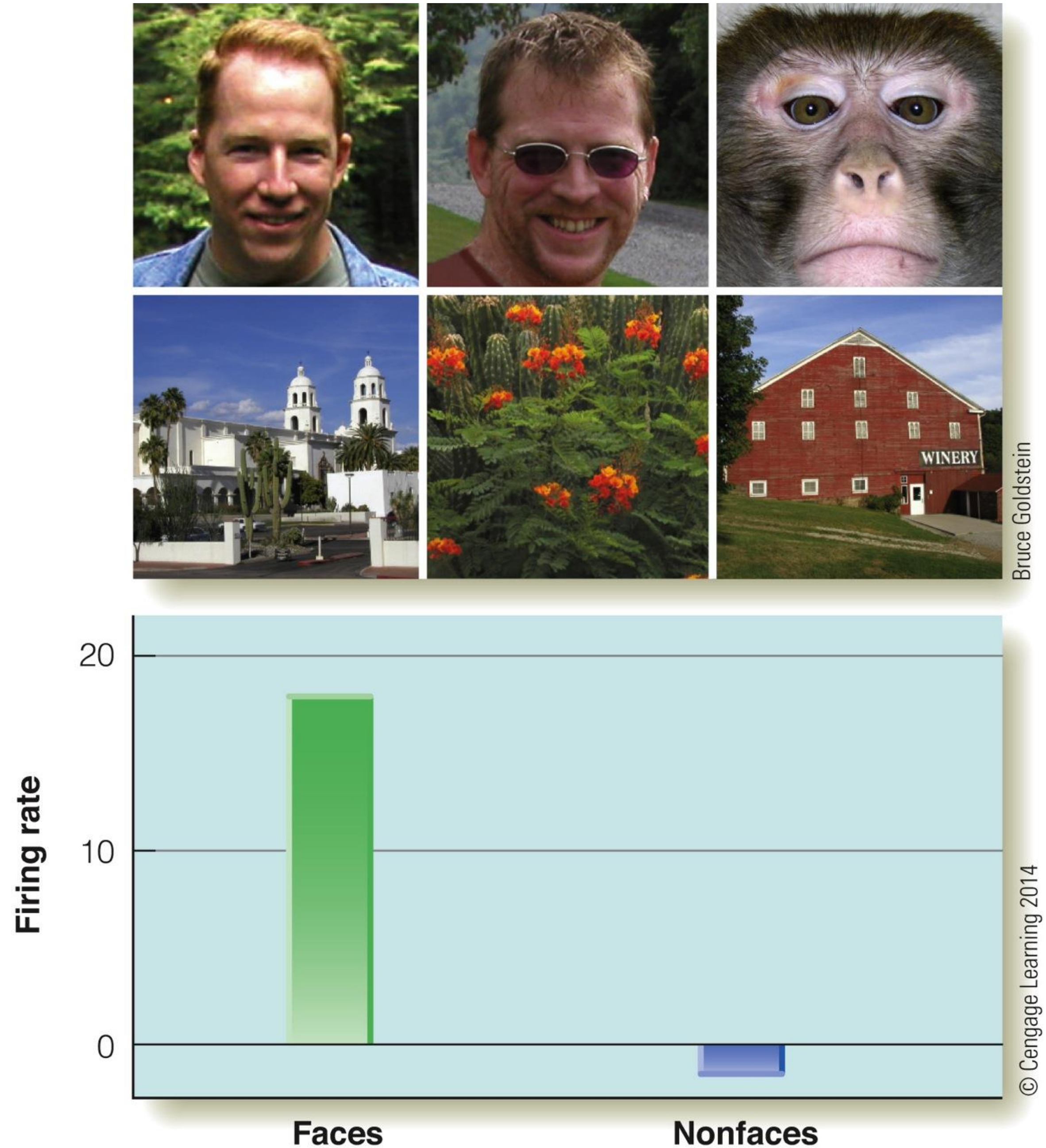


(a) Retina (b) Cortex

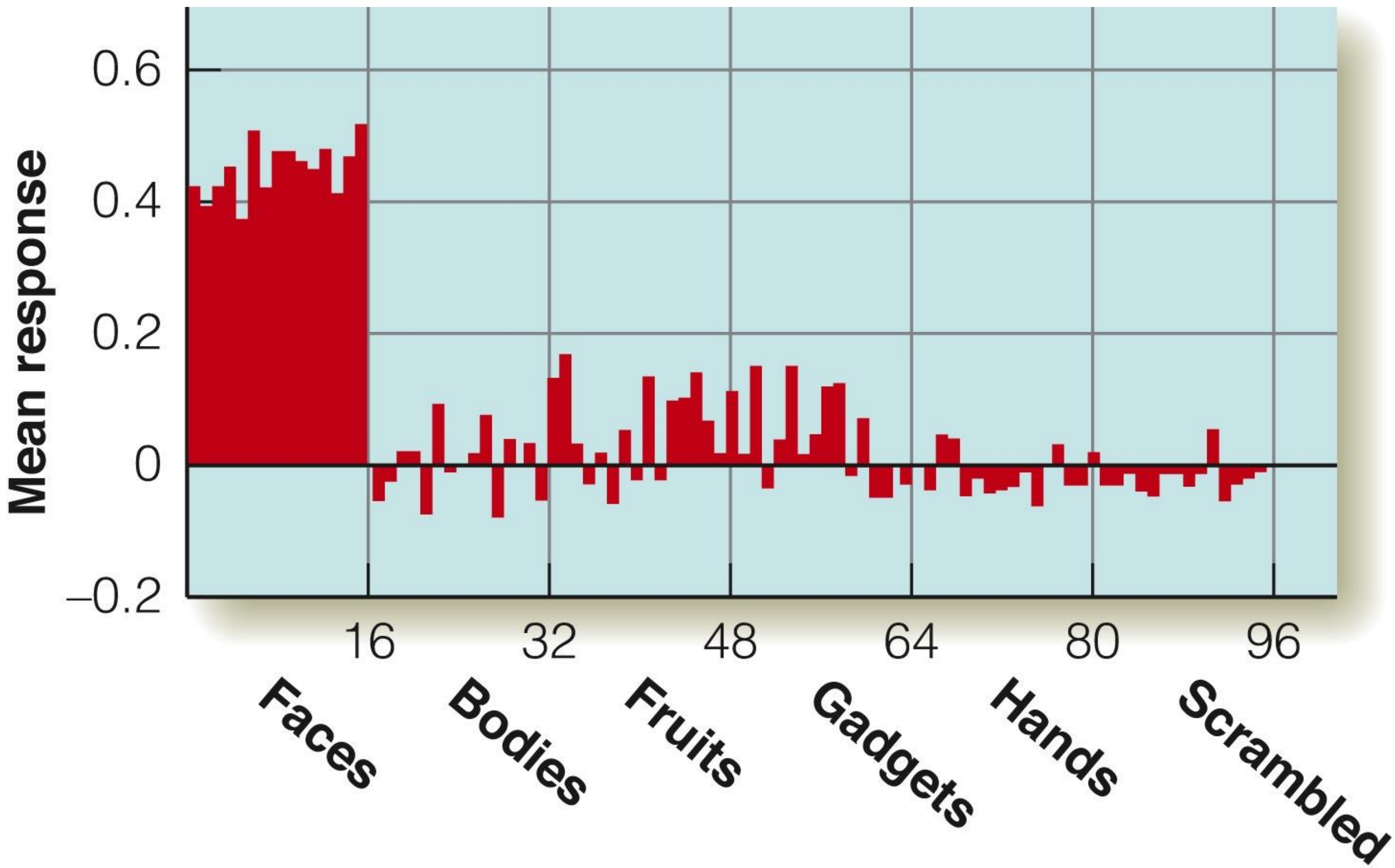

Bruce Goldstein

Higher Level Processing





(Based on data from Rolls & Tovee, 1995.)



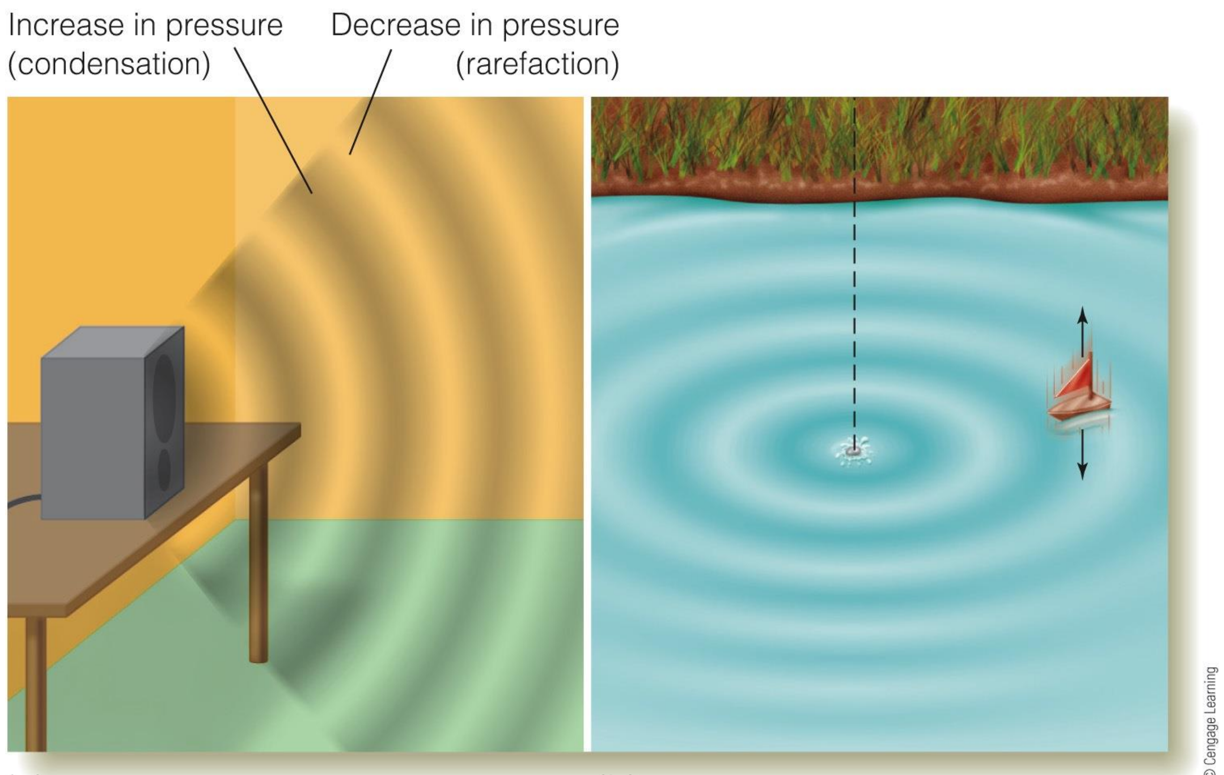
From Tsao, D. Y., Freiwald, W. A., Tootell, R. B., & Livingstone, M. S. (2006). A cortical region consisting entirely of face-selective cells. *Science, 311*, 670–674., Fig. 2b, right. Reprinted with permission from AAAS.

Results of the Tsao et al. (2006) experiment

Hearing & Auditory Perception

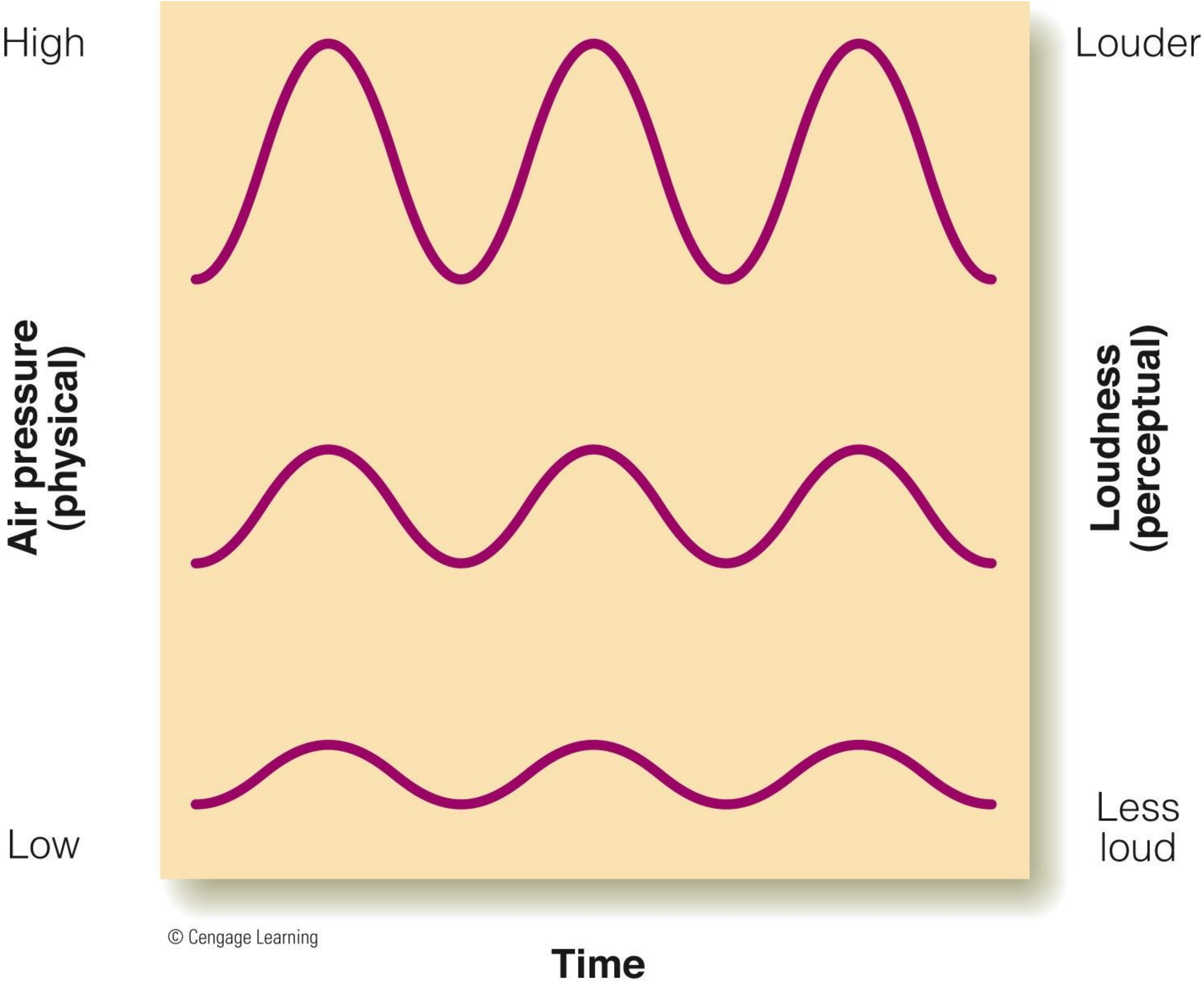
Physical Aspects of Sound

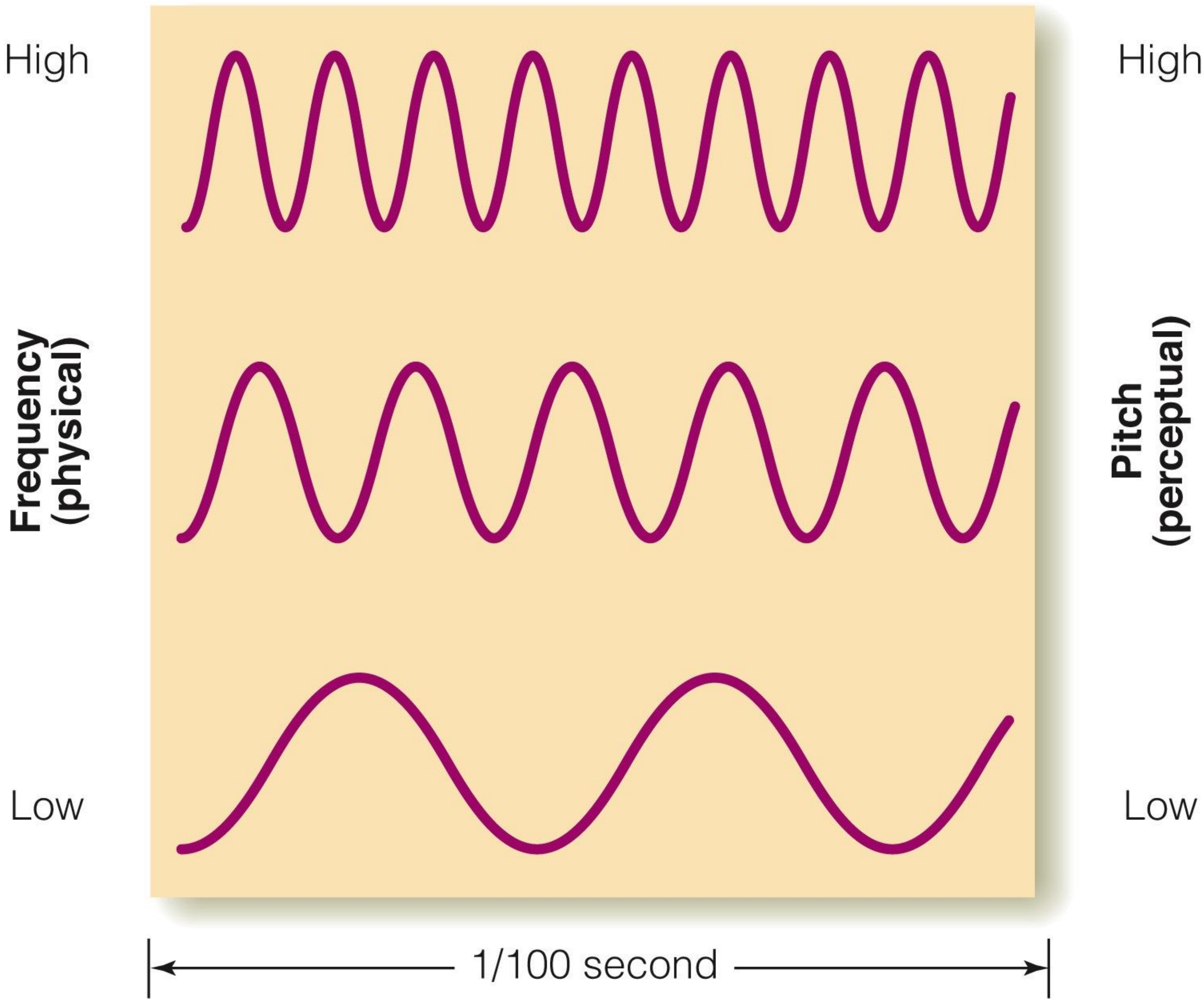
- Two definitions of "sound"
  - Physical definition - sound is *pressure changes* in the air or other medium.
  - Perceptual definition - sound is the *experience* we have when we hear.



Pure Tones

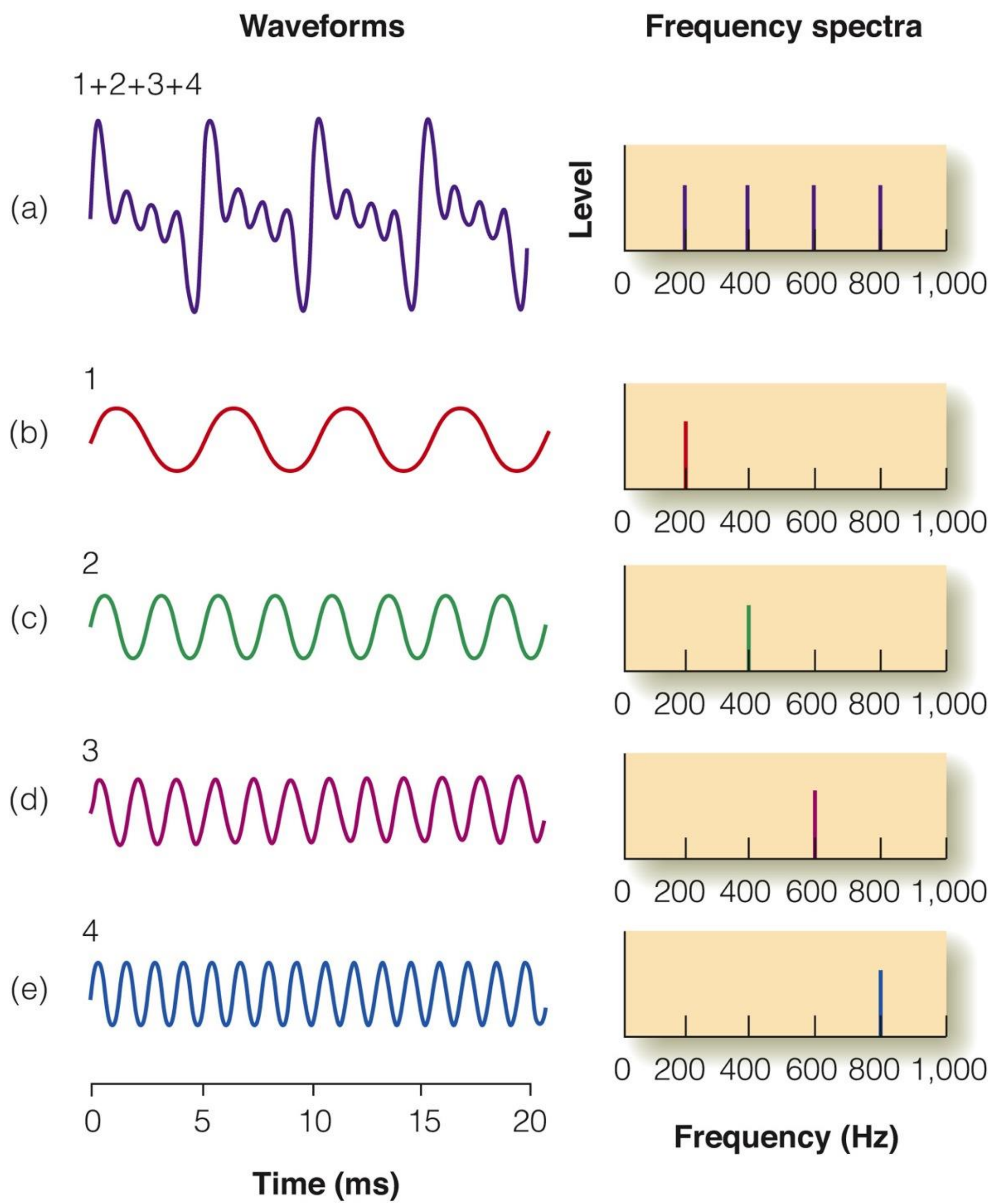
- Pure tone - created by a sine wave
  - Amplitude: difference in pressure between high and low peaks of wave
    - Perception of amplitude is loudness
    - Decibel (dB) is used as the measure of loudness: $dB = 20 \log(p/p_o)$
  - Frequency: number of cycles within a given time period
    - Measured in Hertz (Hz) - cycle per second
    - Perception of pitch is related to frequency.


High
Low
Air pressure (physical)
Louder
Less loud
Loudness (perceptual)
Time


High
High
Frequency (physical)
Pitch (perceptual)
Low
Low
1/100 second

Relative Amplitudes and Decibels for Environmental Sounds

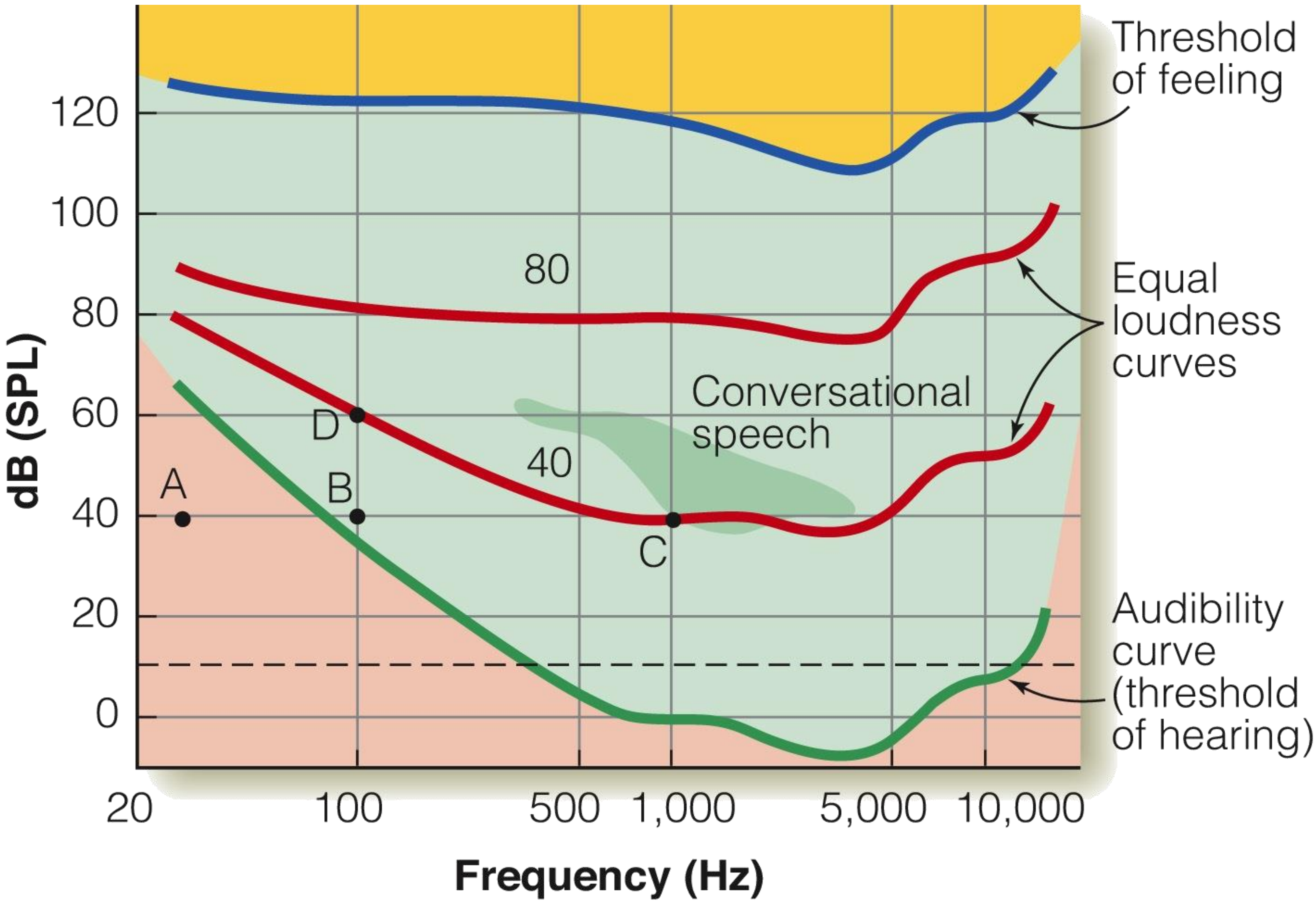
SOUND	RELATIVE AMPLITUDE	DECIBELS (DB)
Barely audible (threshold)	1	0
Leaves rustling	10	20
Quiet residential community	100	40
Average speaking voice	1,000	60
Express subway train	100,000	100
Propeller plane at takeoff	1,000,000	120
Jet engine at takeoff (pain threshold)	10,000,000	140



Adapted from Plack, C. J. (2005). *The sense of hearing.* New York: Psychology Press.

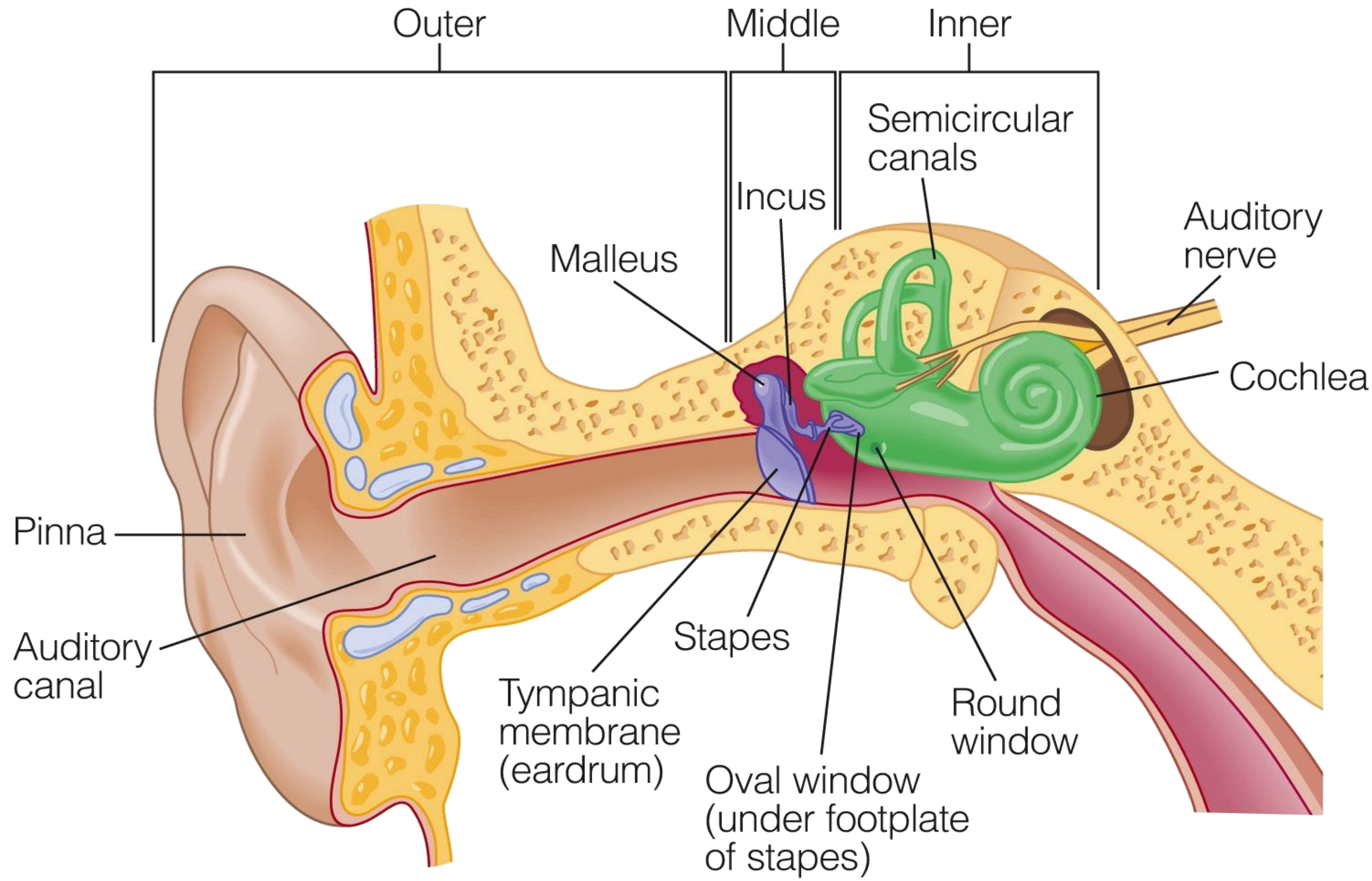
Perceptual Aspects of Sound

- Human hearing range - 20 to 20,000 Hz
- Audibility curve - shows the threshold of hearing in relation to frequency
  - Changes on this curve show that humans are most sensitive to 2,000 to 4,000 Hz.
- Auditory response area - falls between the audibility curve and the threshold for feeling
  - It shows the range of response for human audition.



From Fletcher, H., & Munson, W. A. (1933). Loudness: Its definition, measurement, and calculation. *Journal of the Acoustical Society of America, 5*, 82–108. Reproduced by permission.

The Ear: Sound Transduction Organ



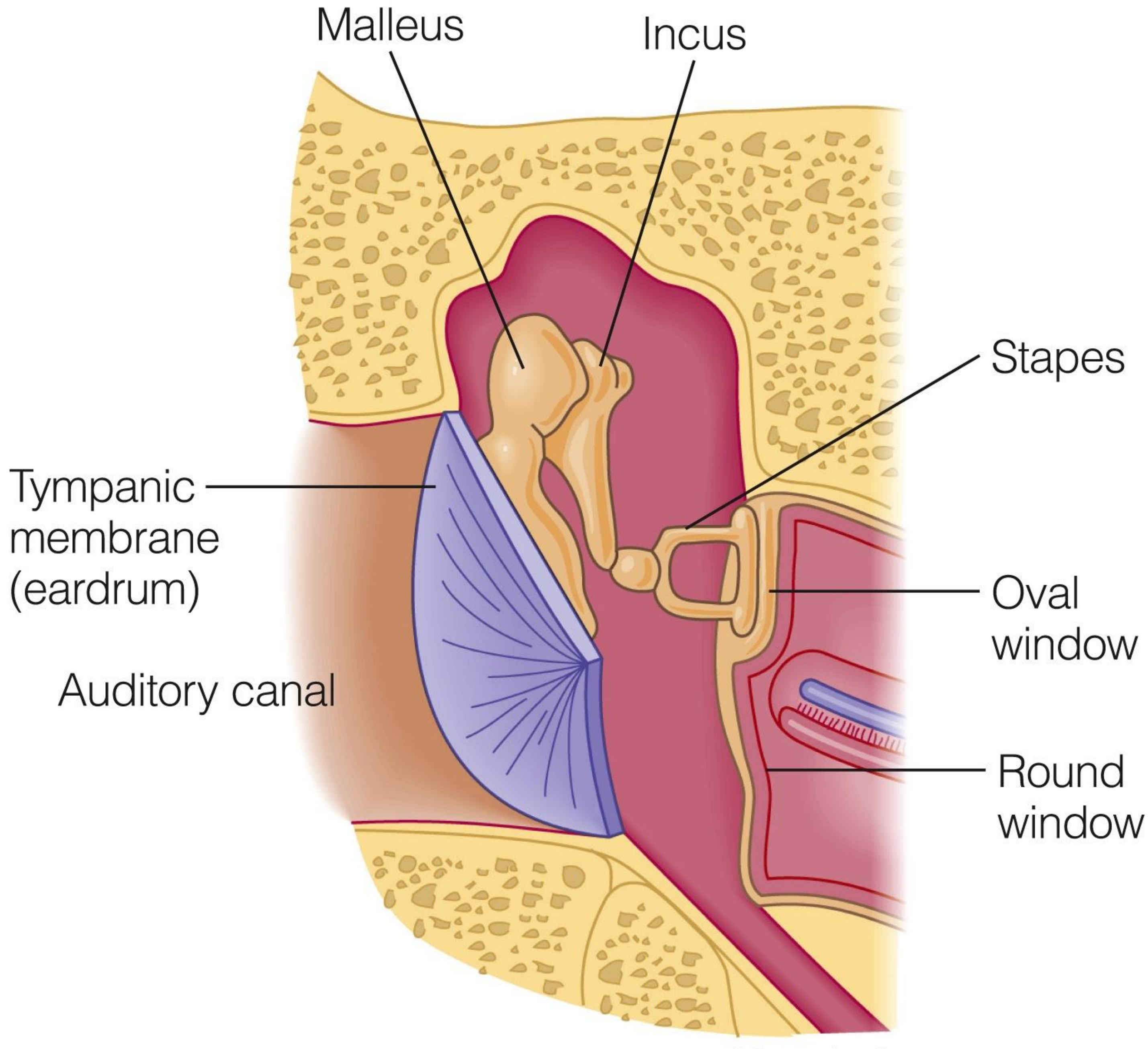
From Lindsay, P. H., & Norman, D. A. (1977). *Human information processing* (2nd ed.). New York: Academic Press. Redrawn by permission.

From Pressure Changes to Electricity

- Outer ear - pinna and auditory canal
  - Pinna helps with sound location.
  - Auditory canal - tube-like 3 cm long structure
    - It protects the tympanic membrane at the end of the canal.
    - The resonant frequency of the canal amplifies frequencies between 1,000 and 5,000 Hz.

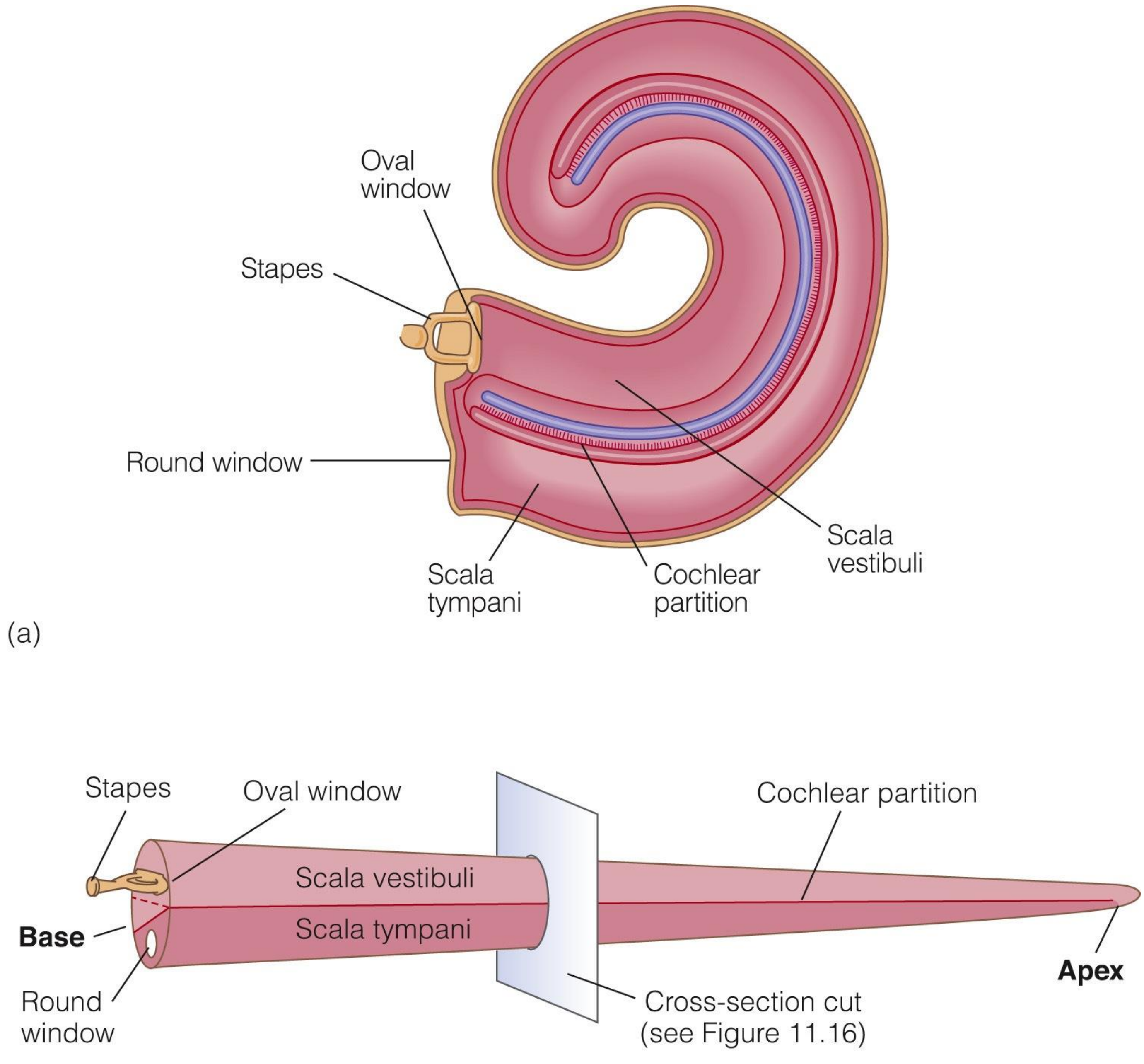
From Pressure Changes to Electricity - continued

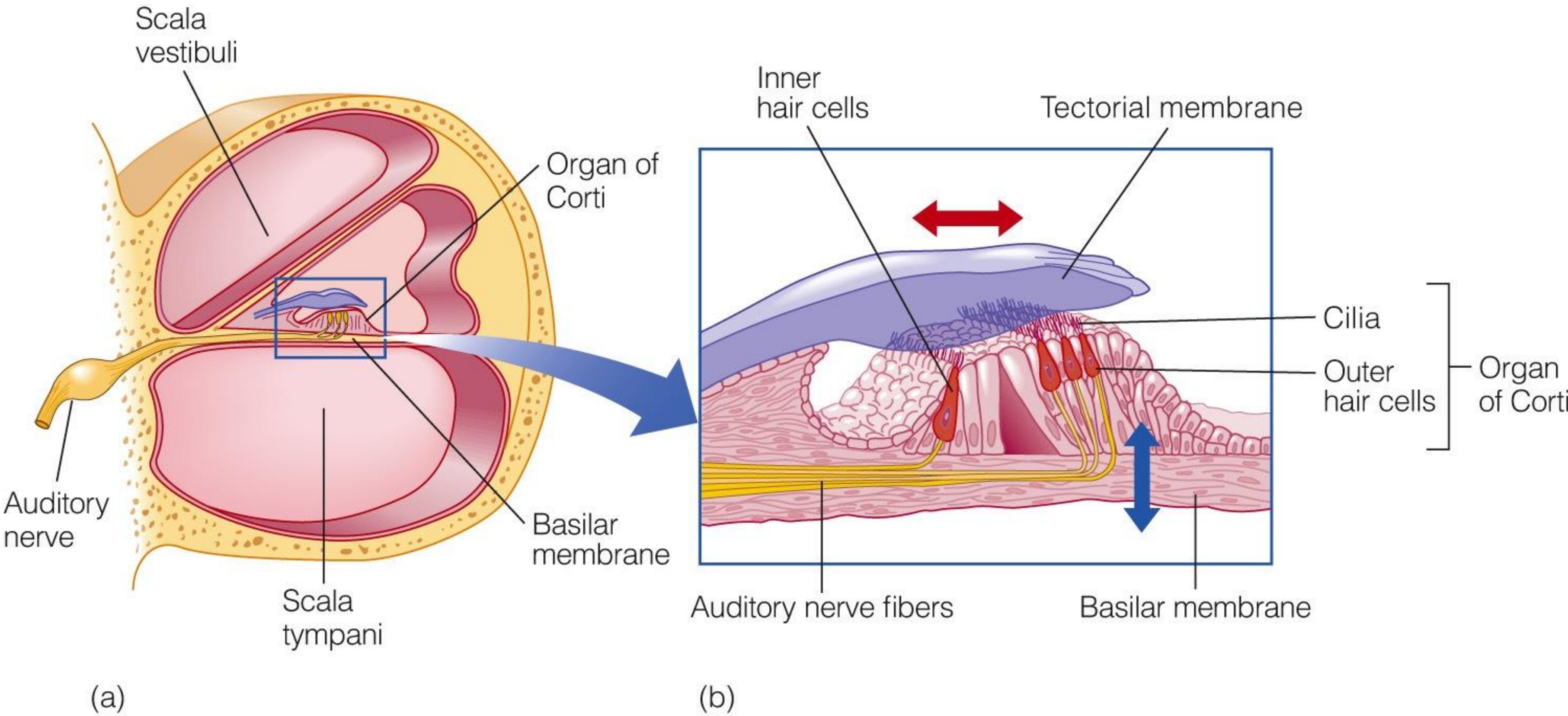
- Middle ear
  - Two cubic centimeter cavity separating inner from outer ear
  - It contains the three ossicles
    - Malleus - moves due to the vibration of the tympanic membrane
    - Incus - transmits vibrations of malleus
    - Stapes - transmit vibrations of incus to the inner ear via the oval window of the cochlea


Malleus
Incus
Stapes
Tympanic membrane (eardrum)
Oval window
Auditory canal
Round window
© Cengage Learning

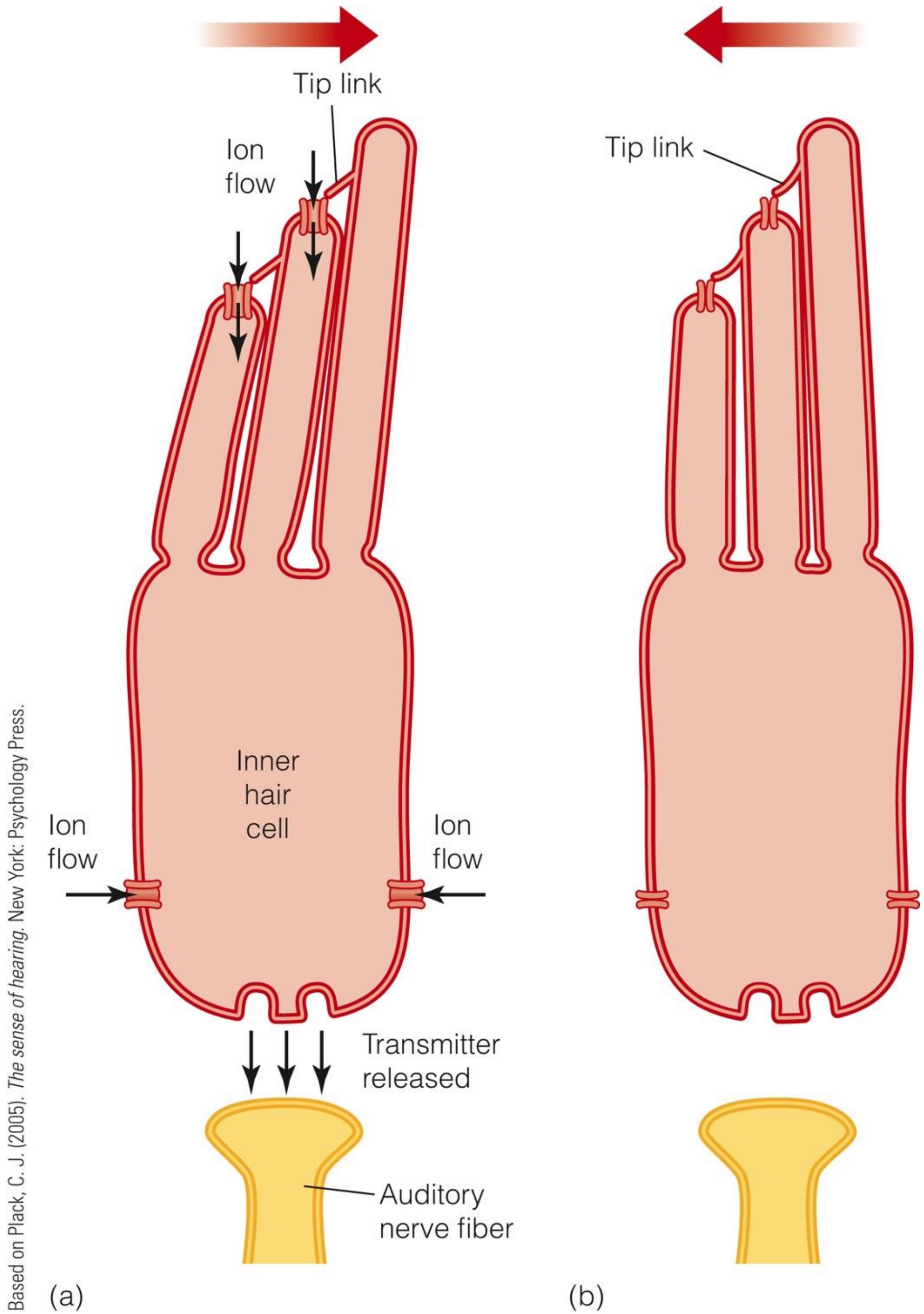
From Pressure Changes to Electricity - continued

- Inner ear
  - Main structure is the cochlea
    - Fluid-filled snail-like structure (35 mm long) set into vibration by the stapes
    - Divided into the scala vestibuli and scala tympani by the cochlear partition
    - Cochlear partition extends from the base (stapes end) to the apex (far end)
    - Organ of Corti contained by the cochlear partition


Oval
window
Stapes
Round window
Scala
tympani
Cochlear
partition
Scala
vestibuli
(a)
Stapes
Oval window
Cochlear partition
Scala vestibuli
Scala tympani
Base
Round
window
Apex
Cross-section cut
(see Figure 11.16)
(b)

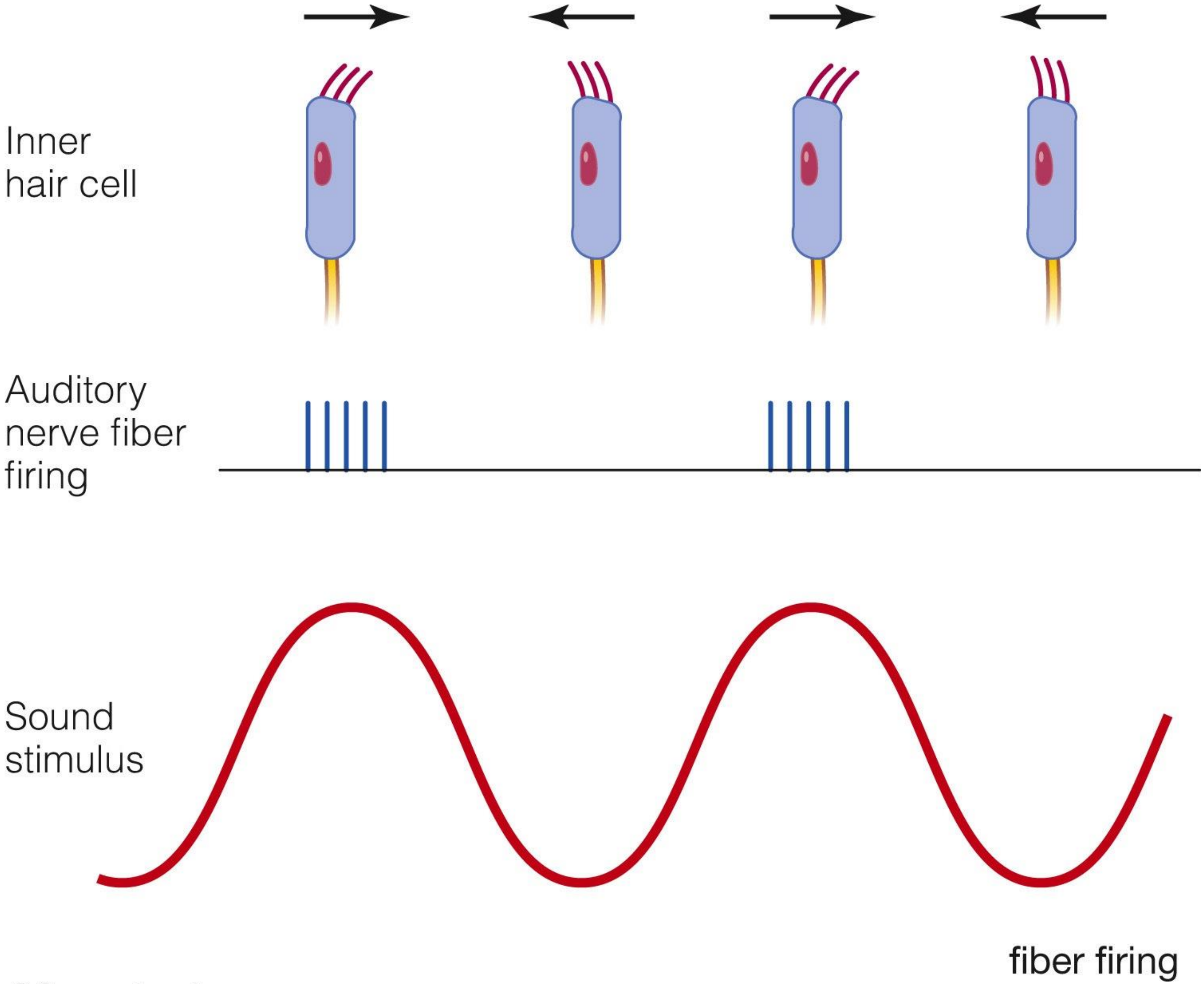


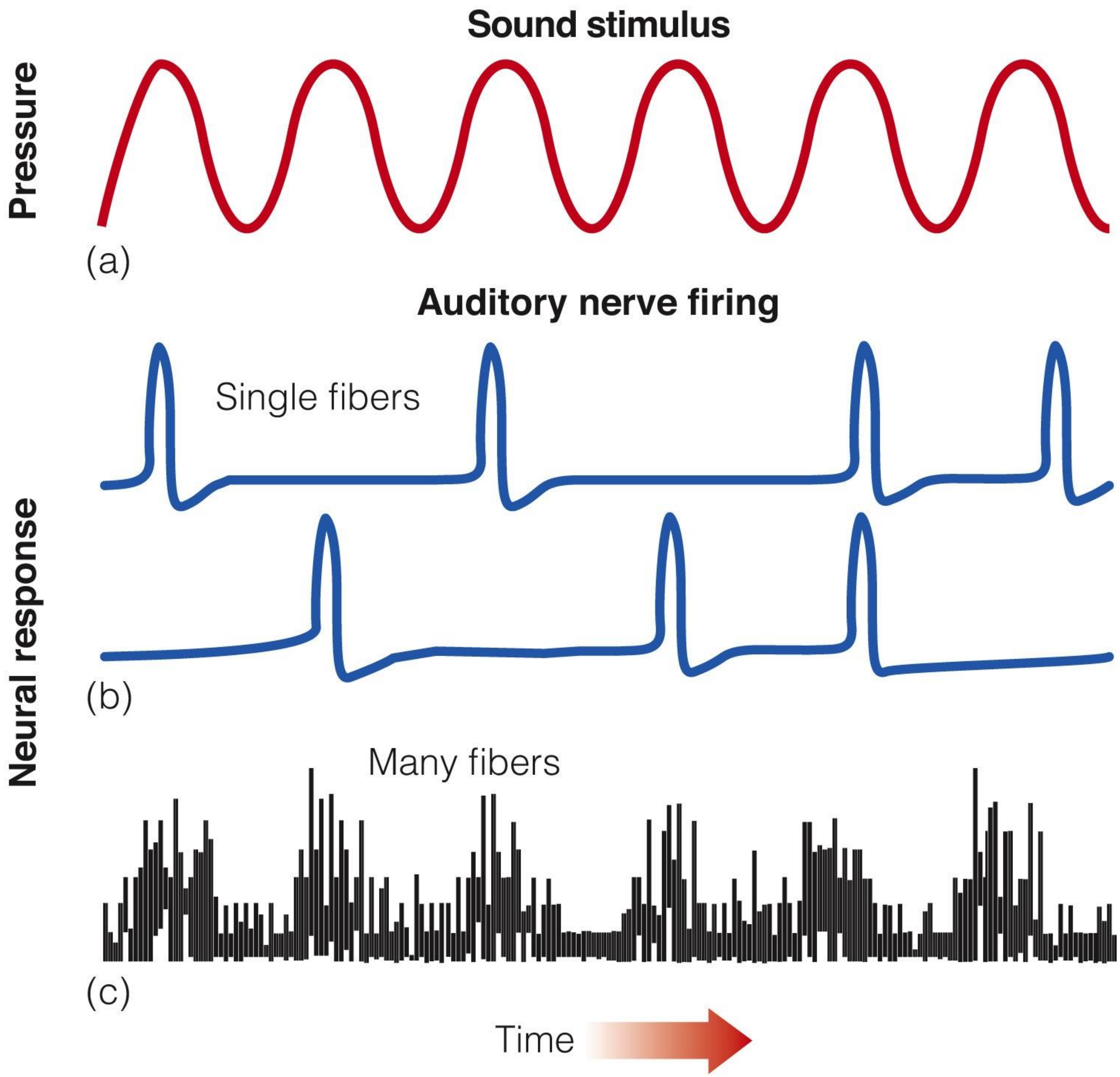
Adapted from Denes, P. B., & Pinson, E.N. (1993). *The speech chain* (2nd ed.). NewYork: Freeman.



Transduction:

- Cilia bend in response to movement of organ of Corti and the tectorial membrane
- Movement in one direction opens ion channels
- Movement in the other direction closes the channels
- This causes bursts of electrical signals


Inner
hair cell
Auditory
nerve fiber
firing
Sound
stimulus
fiber firing

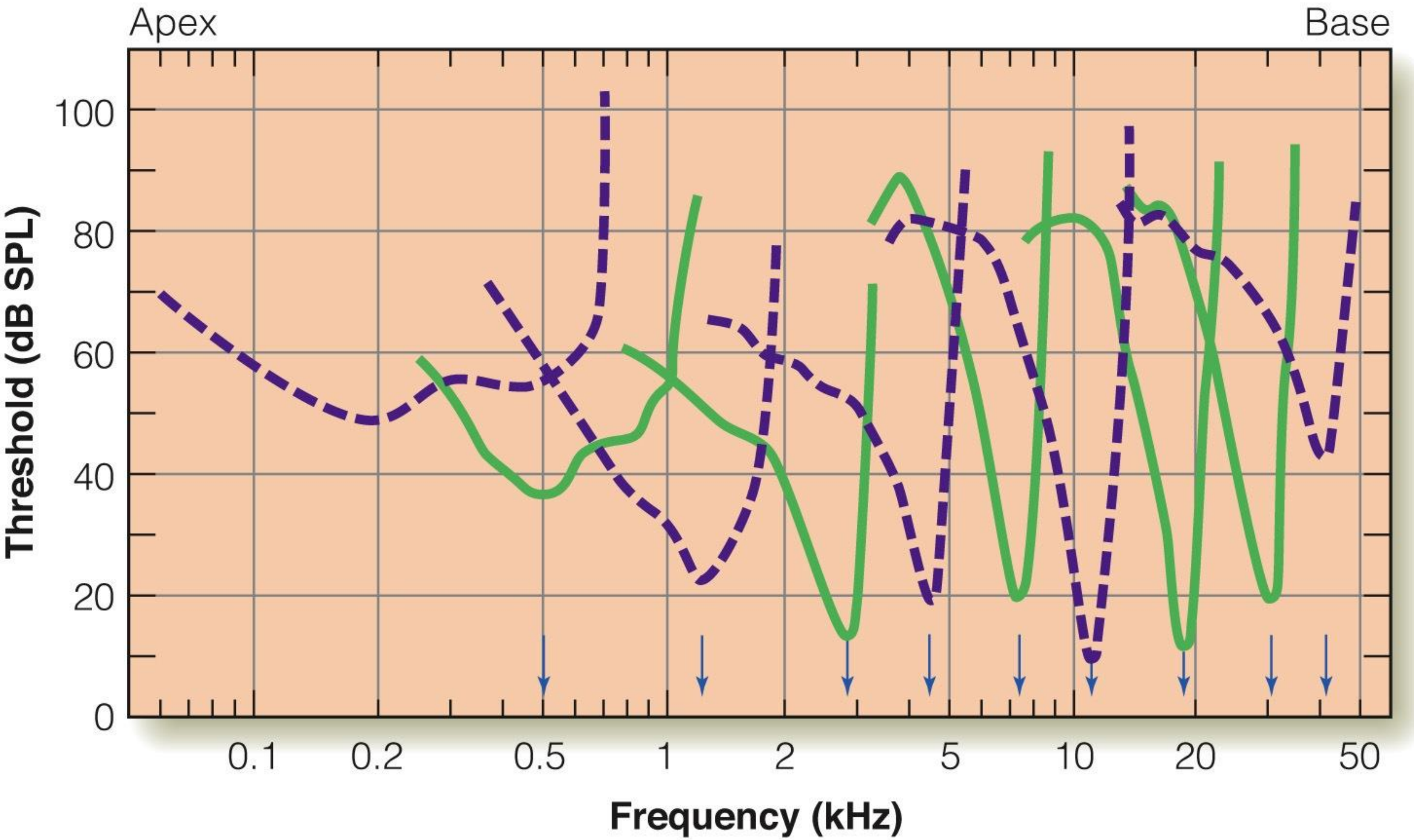


Based on Plack, C. J. (2005). *The sense of hearing*. New York: Psychology Press.

Békésy's Place Theory of Hearing (1960)

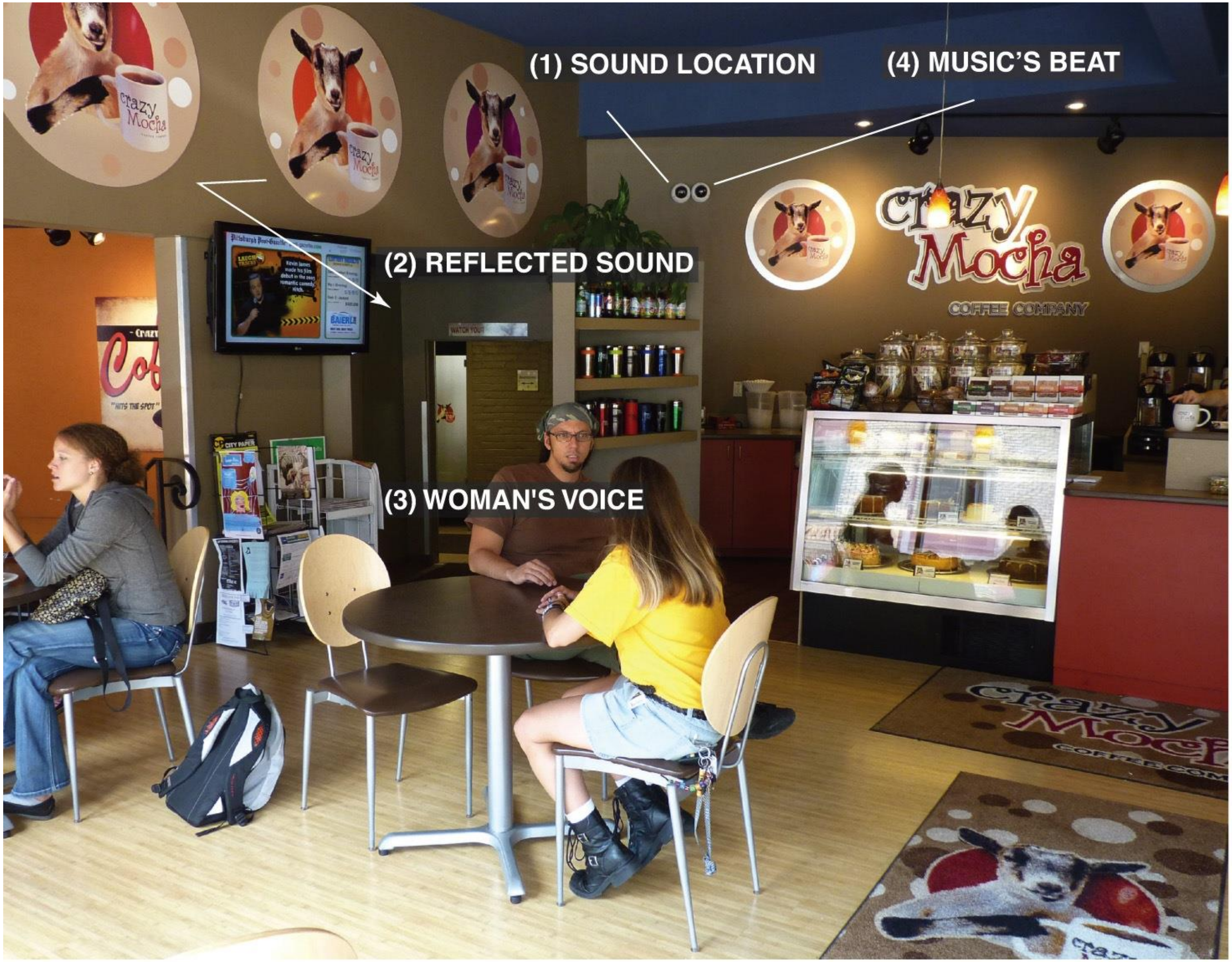
- Frequency of sound is indicated by the place on the organ of Corti that has the highest firing rate.
- Békésy determined this in two ways:
  - Direct observation of the basilar membrane in cadavers.
  - Building a model of the cochlea using the physical properties of the basilar membrane.

Evidence for Place Theory

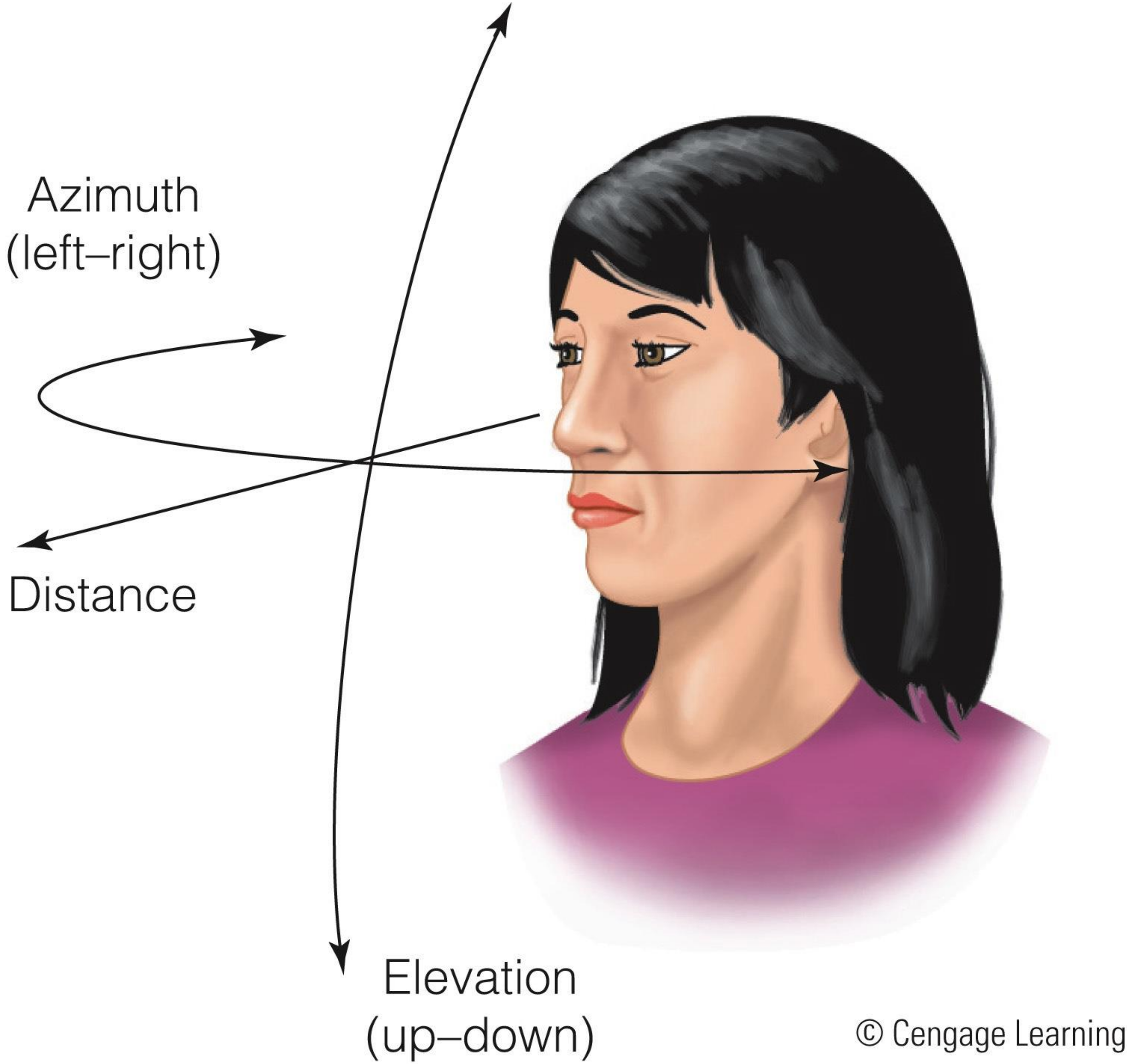


From Palmer, A. R. Physiology of the cochlear nerve and cochlear nucleus. *British Medical Bulletin on Hearing, 43,* 1987, 838–855, by permission of Oxford University Press.

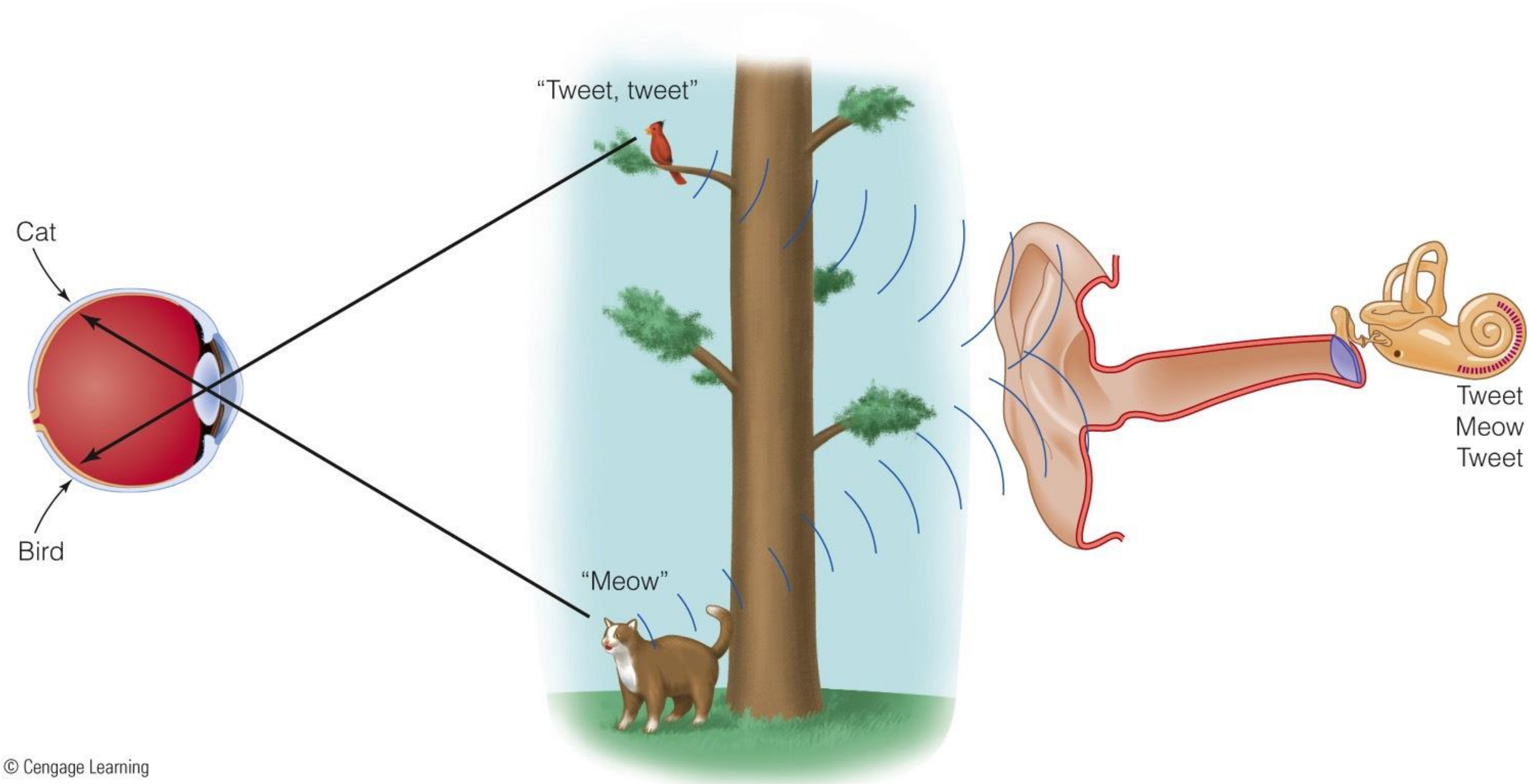
Auditory Localization and Organization



Bruce Goldstein

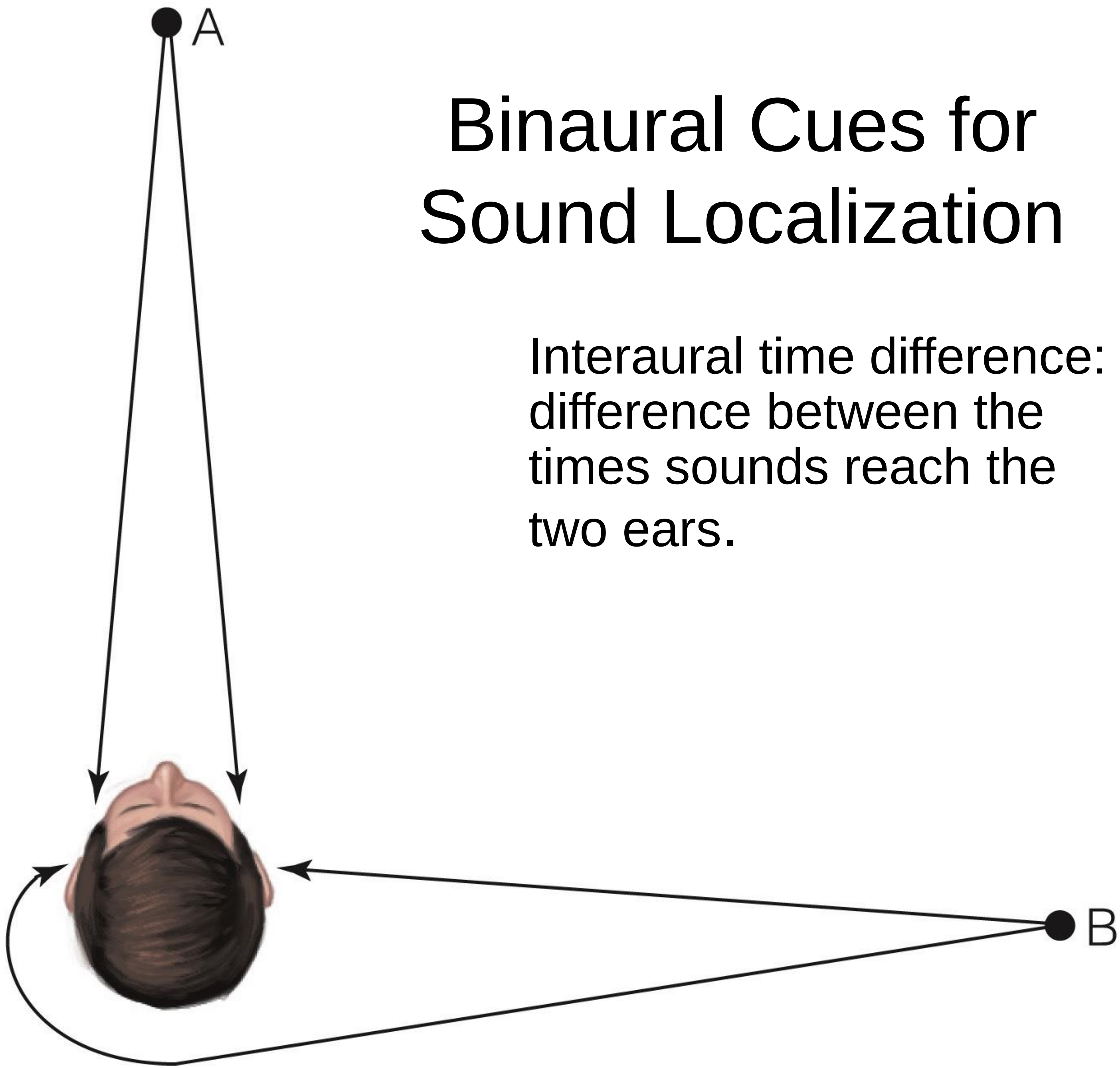


Azimuth
(left–right)
Distance
Elevation
(up–down)
© Cengage Learning

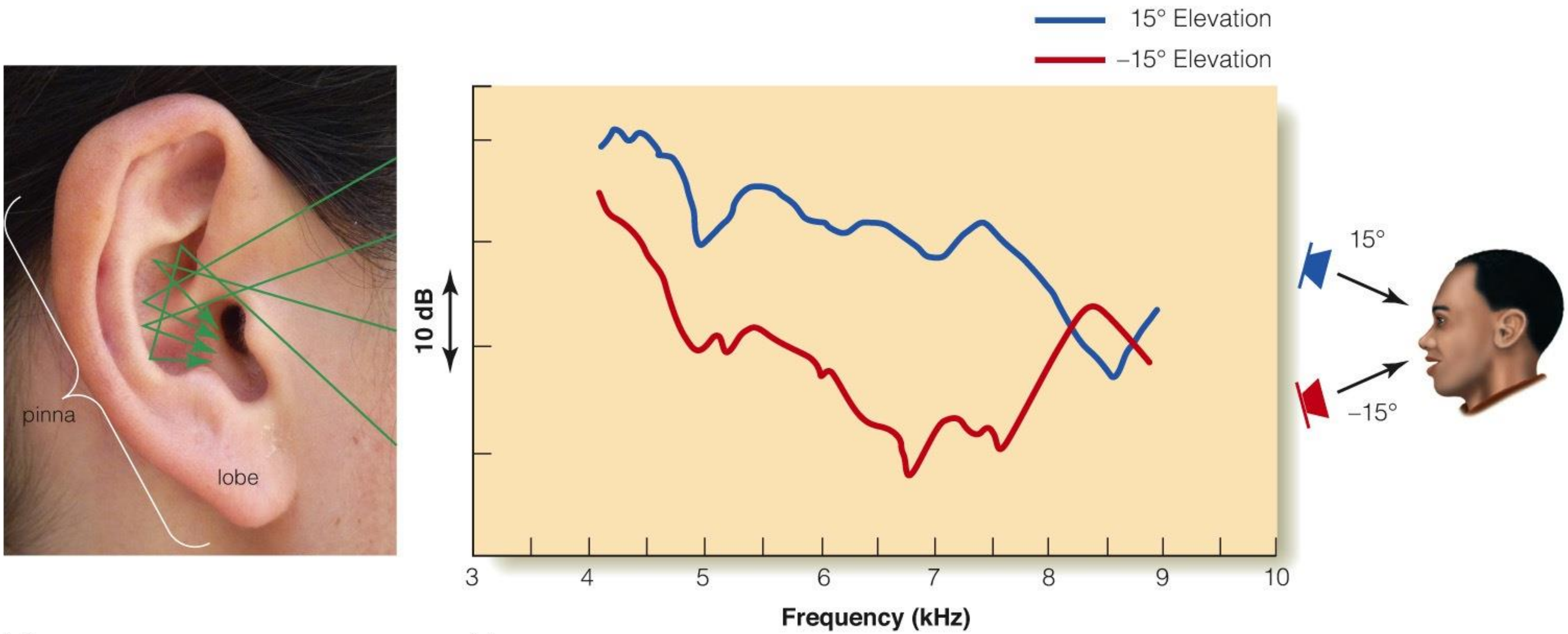

Cat
Bird
"Tweet, tweet"
"Meow"
Tweet
Meow
Tweet

Binaural Cues for Sound Localization

Interaural time difference: difference between the times sounds reach the two ears.



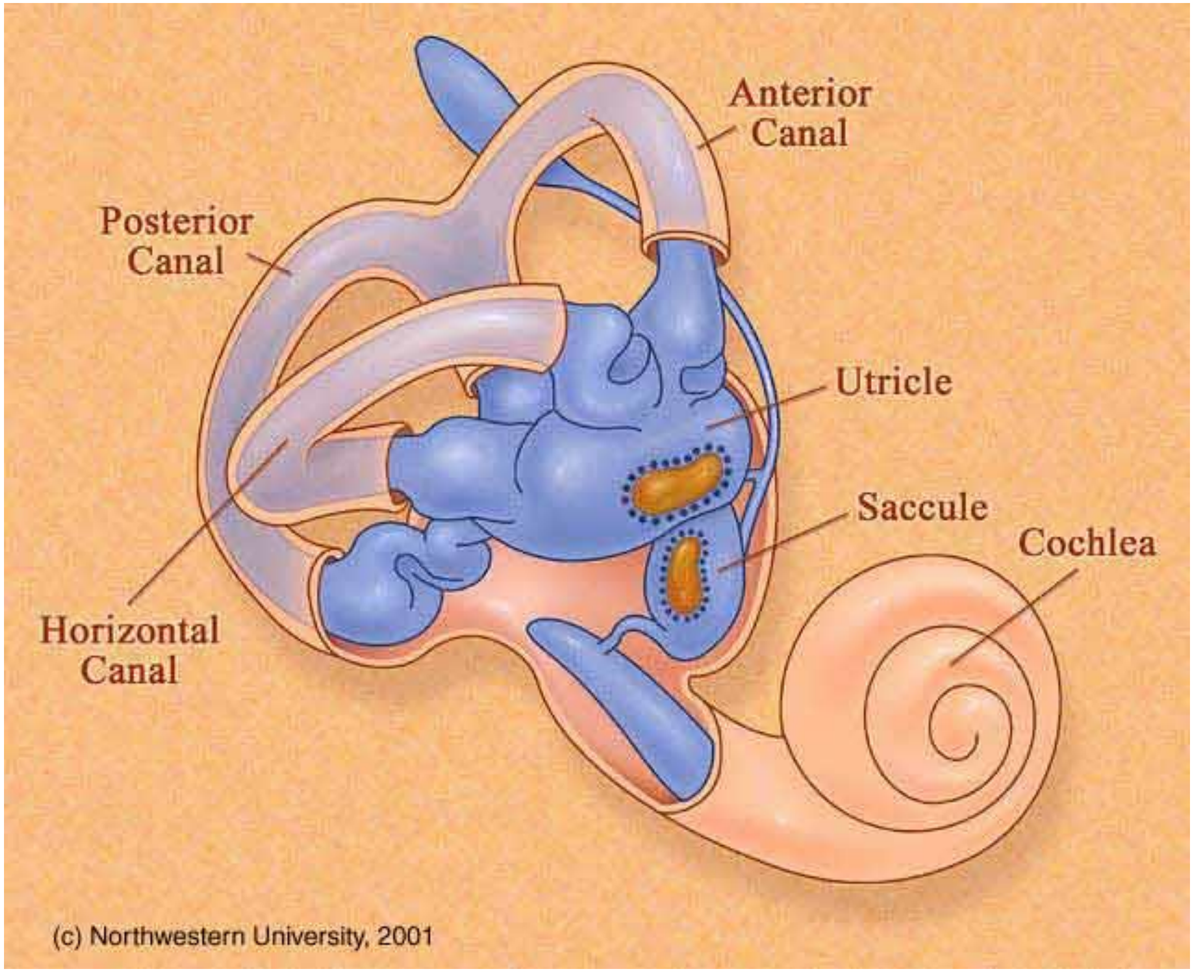
Monaural Cue for Sound Location

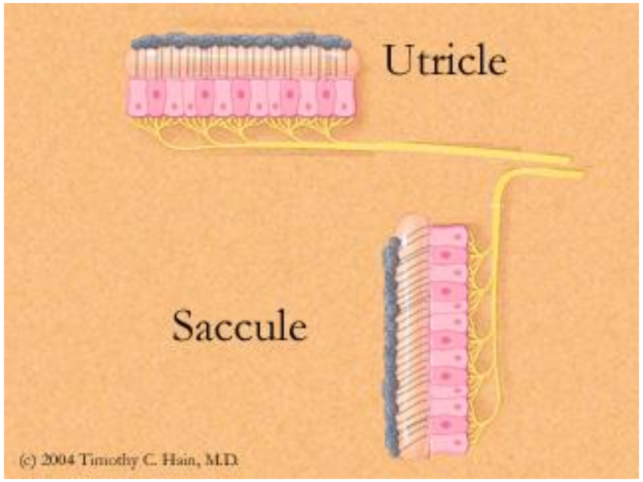


(a) (b)

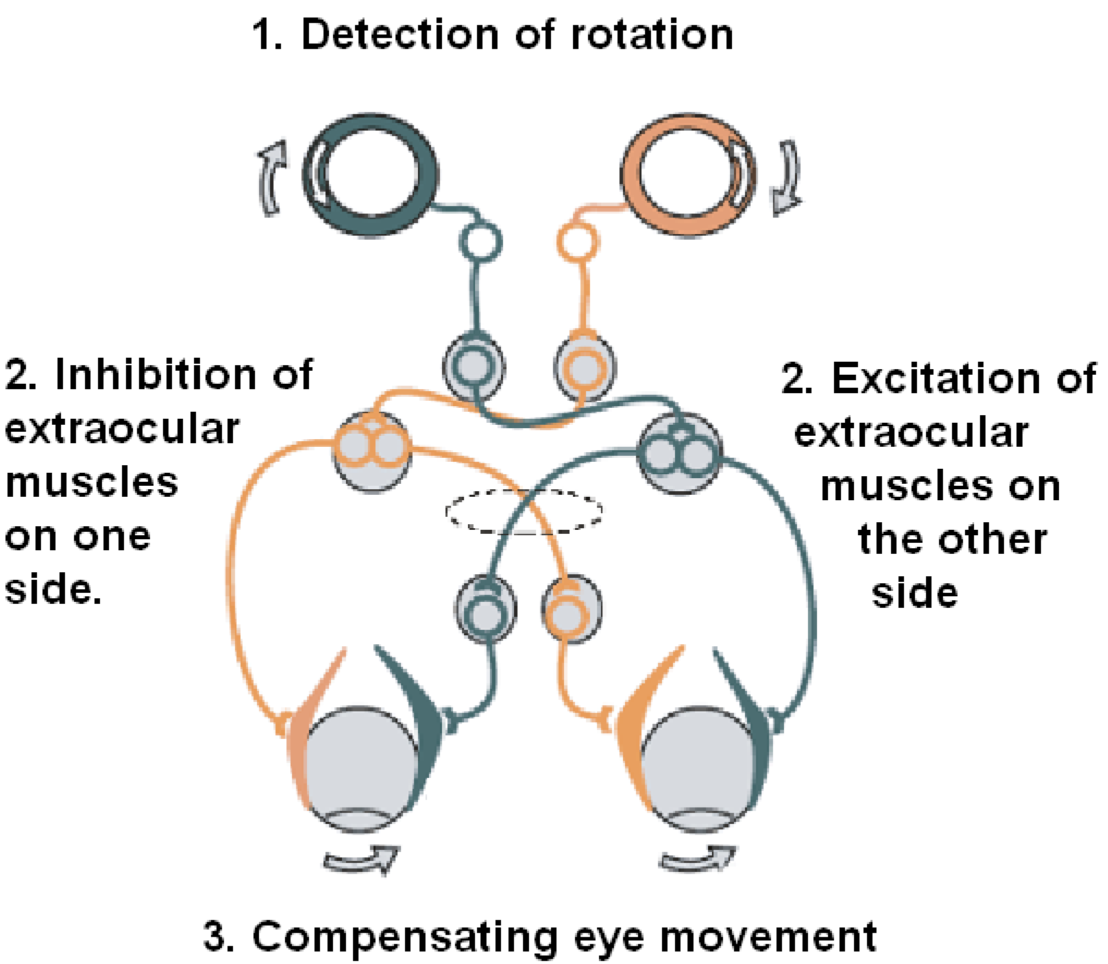
Adapted from Plack, C. J. (2005). *The sense of hearing*, Figure 9.11. New York: Psychology Press. Ear photo by Bruce Goldstein.

Human Vestibular System





Vestibulo-Ocular Reflex

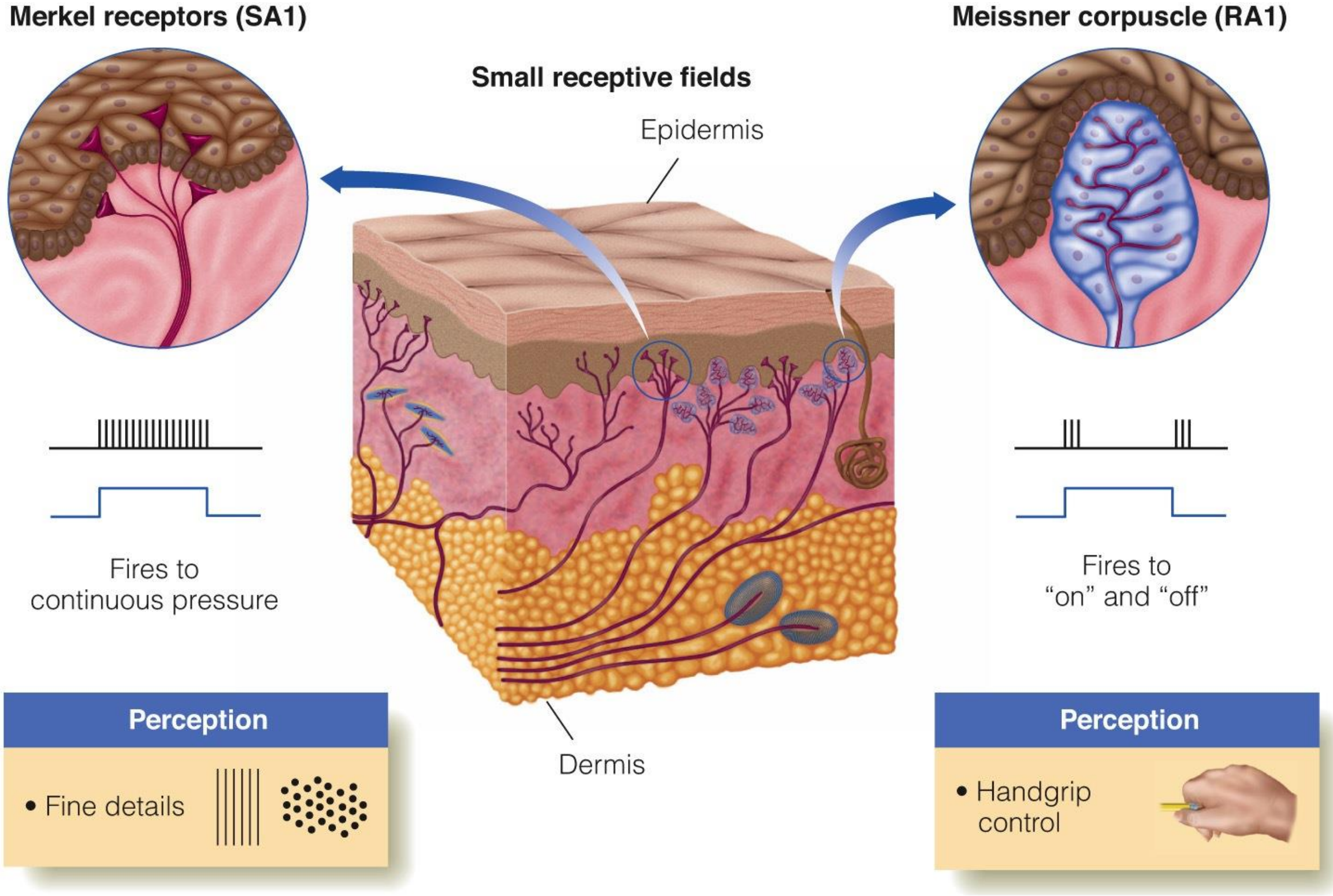


Touch & Tactile Perception

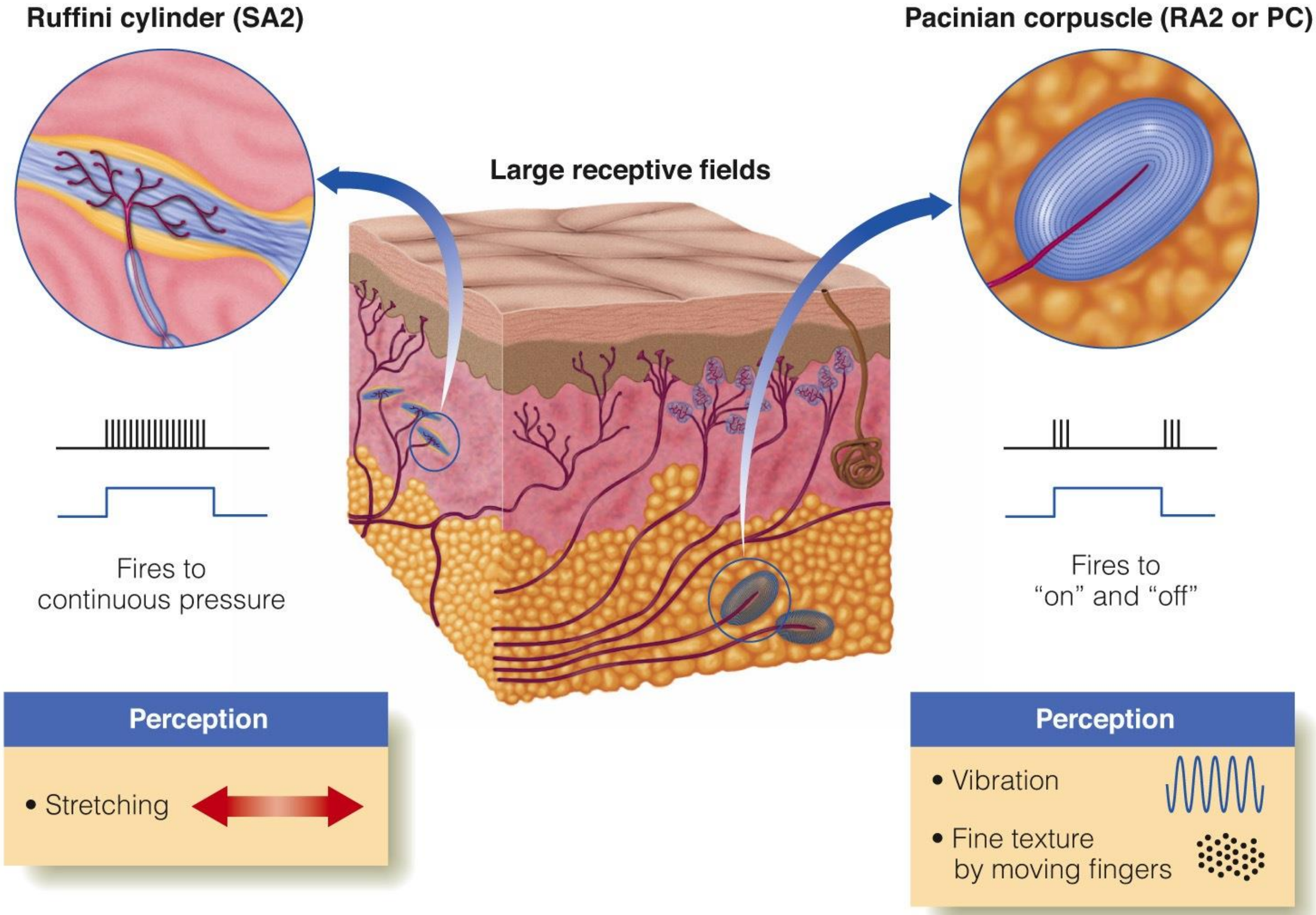
Somatosensory System

- There are three parts
  - Cutaneous senses - perception of touch and pain from stimulation of the skin
  - Proprioception - ability to sense position of the body and limbs
  - Kinesthesis - ability to sense movement of body and limbs

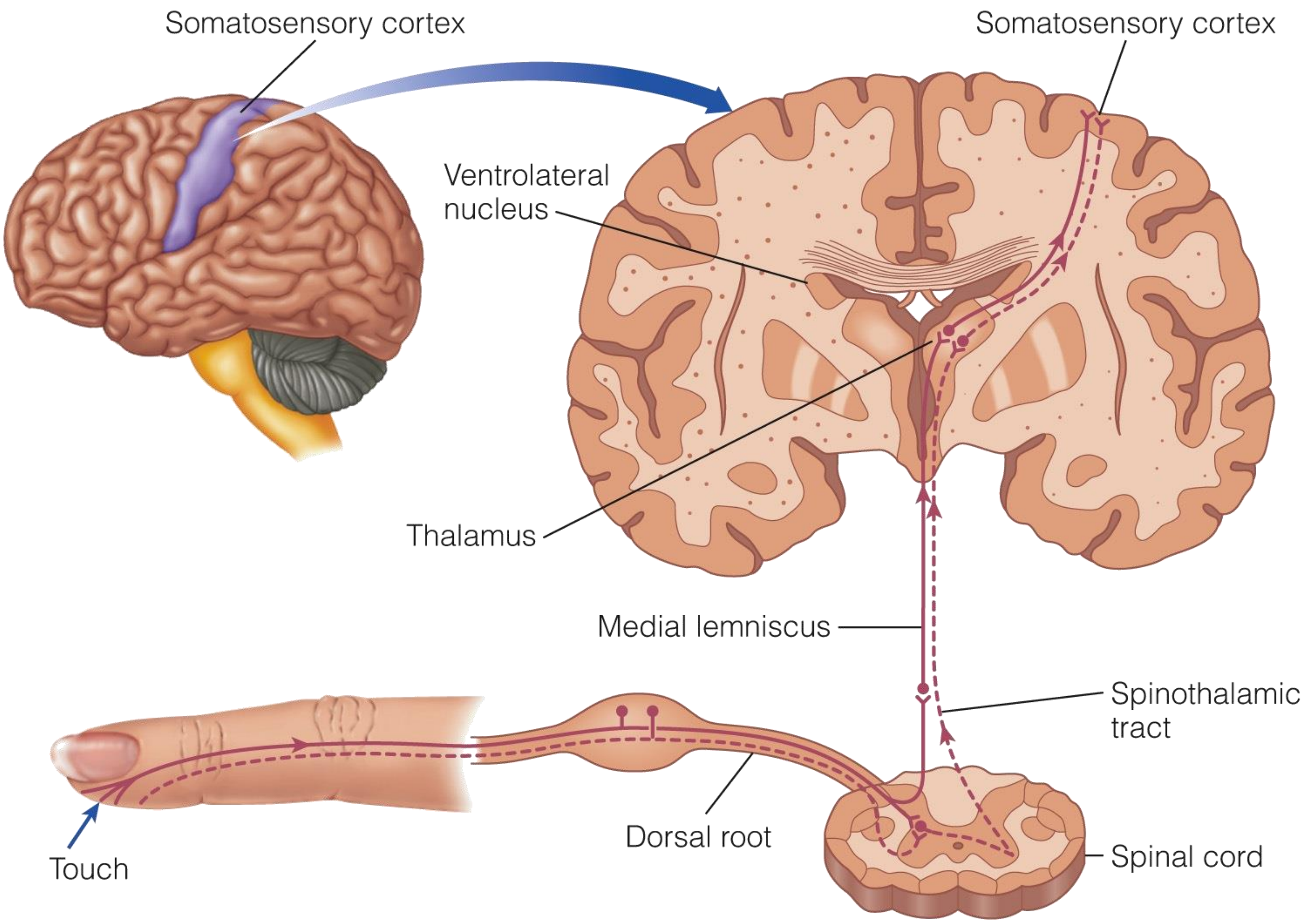
Mechanoreceptors



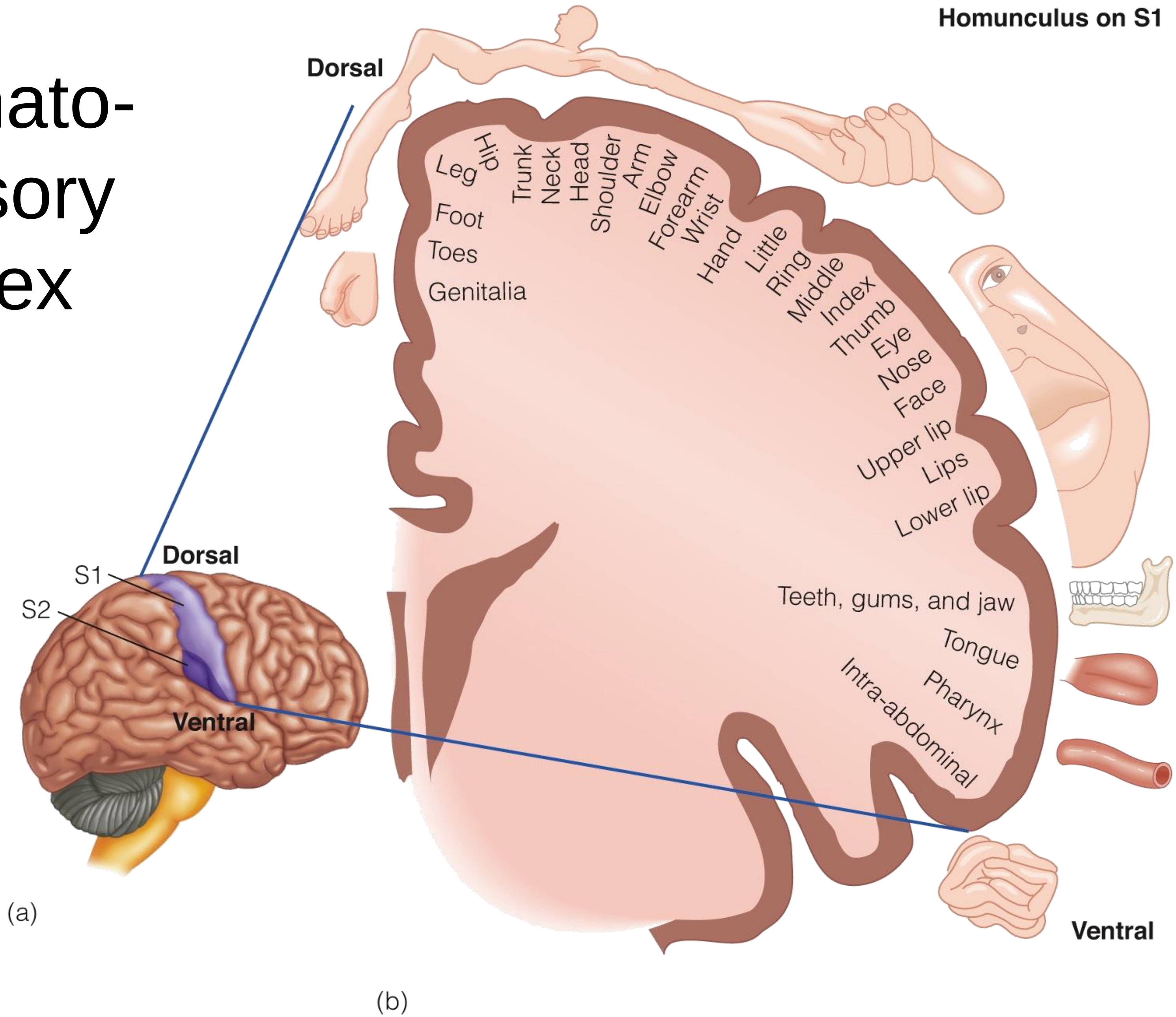
Mechanoreceptors



Pathways from Skin to Cortex



Somato-sensory Cortex

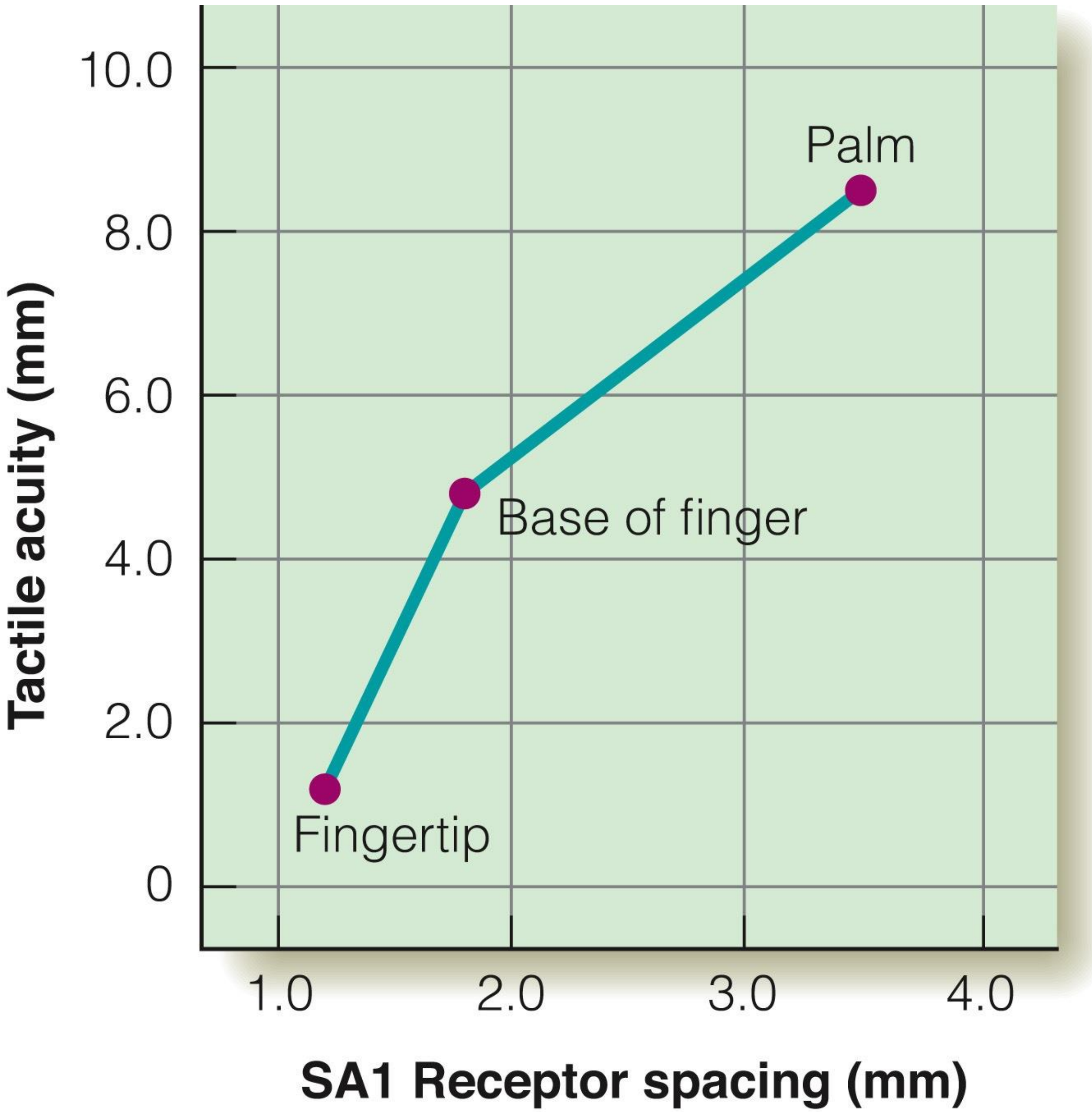


Adapted from Penfield, W., & Rasmussen, T. (1950). *The cerebral cortex of man.* New York: Macmillan.

Receptor Mechanisms for Tactile Acuity

- There is a high density of Merkel receptors in the fingertips.

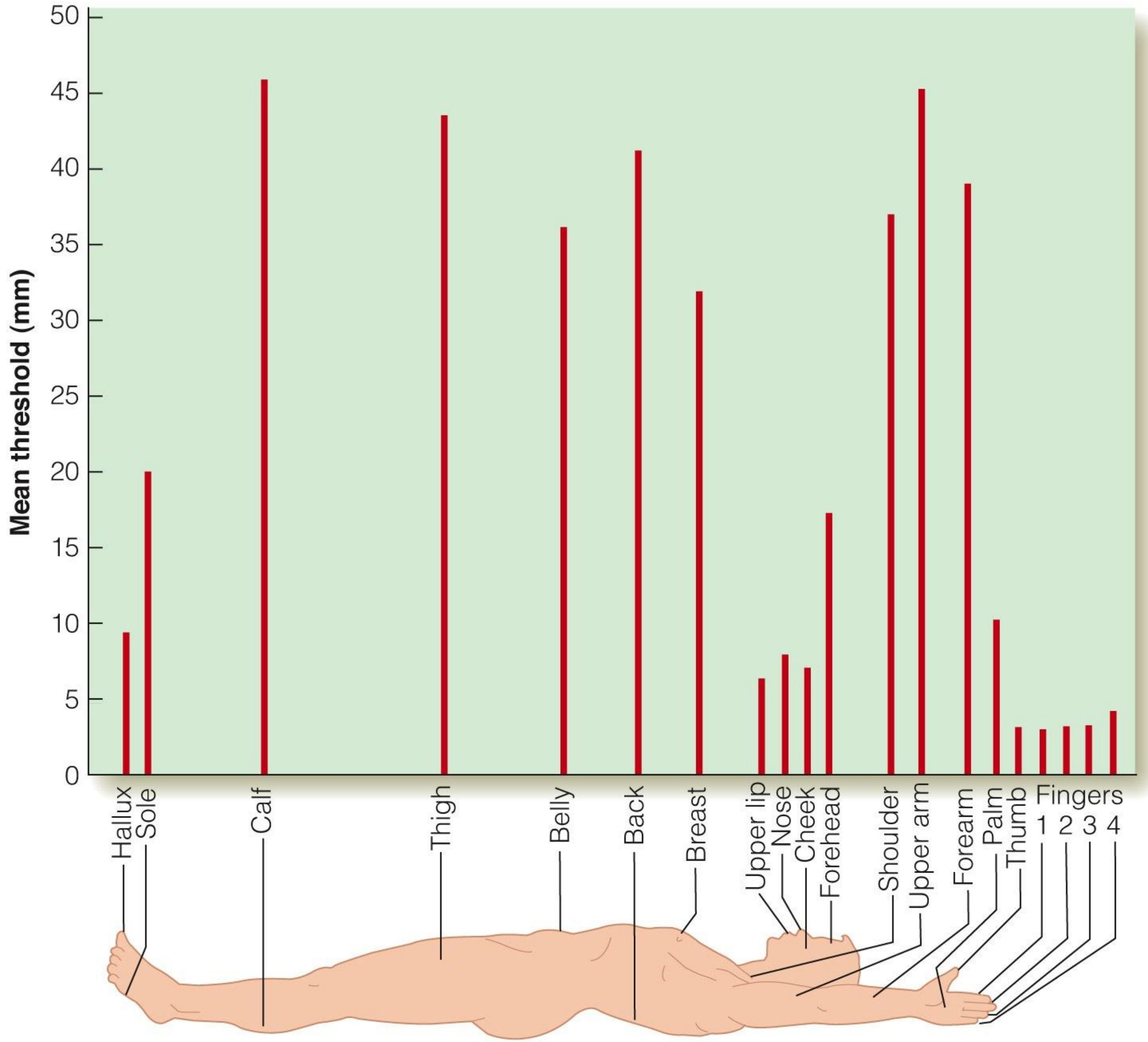
- Merkel receptors are densely packed on the fingertips - similar to cones in the fovea.



From Craig, J. C., & Lyle, K. B. (2002). A correction and a comment on Craig and Lyle (2001). *Perception & Psychophysics, 64*, 504–506.

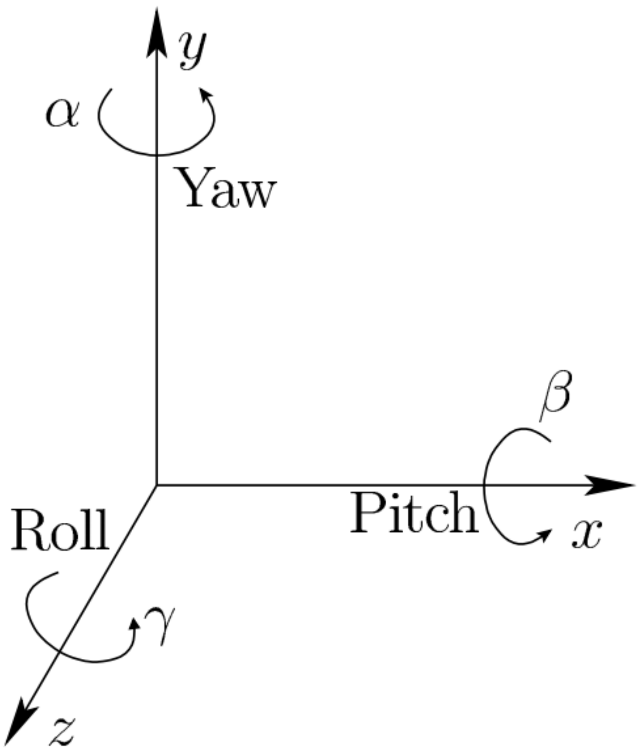
Cortical Mechanisms for Tactile Acuity

- Body areas with high acuity have larger areas of cortical tissue devoted to them.
- This parallels the "magnification factor" seen in the visual cortex for the cones in the fovea.
- Areas with higher acuity also have smaller receptive fields on the skin.


Mean threshold (mm)
0
5
10
15
20
25
30
35
40
45
50
Hallux
Sole
Calf
Thigh
Belly
Back
Breast
Upper lip
Nose
Cheek
Forehead
Shoulder
Upper arm
Forearm
Palm
Thumb
Fingers
1
2
3
4

Backup Slides

3D Rotation: Yaw, Pitch and Roll



$$R_y(\alpha) = \begin{bmatrix} \cos\alpha & 0 & \sin\alpha \\ 0 & 1 & 0 \\ -\sin\alpha & 0 & \cos\alpha \end{bmatrix}$$ Yaw

$$R_x(\beta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\beta & -\sin\beta \\ 0 & \sin\beta & \cos\beta \end{bmatrix}$$ Pitch

$$R_z(\gamma) = \begin{bmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix}$$ Roll

Combining rotations: $R(\alpha, \beta, \gamma) = R_y(\alpha) R_x(\beta) R_z(\gamma)$

3D Rotation and 3D Translation

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = R \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} x_t \\ y_t \\ z_t \end{bmatrix}$$

Can the rotation and translation be combined into a single matrix?

Not with a single 3X3 matrix…

Homogeneous Transformation Matrix

$$T_{rb} = \begin{bmatrix} \begin{array}{|ccc|} \hline & & \\ & R & \\ & & \\ \hline \end{array} & \begin{array}{c} x_t \\ y_t \\ z_t \end{array} \\ \begin{array}{ccc} 0 & 0 & 0 \end{array} & 1 \end{bmatrix}$$

$$\begin{bmatrix} \begin{array}{|ccc|} \hline & & \\ & R & \\ & & \\ \hline \end{array} & \begin{array}{c} x_t \\ y_t \\ z_t \end{array} \\ \begin{array}{ccc} 0 & 0 & 0 \end{array} & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix}$$

Note:

rb denotes rigid body

The sequence in this matrix is rotation followed by translation

S. Lavalle

How Would We Inverse (“Undo”) This?

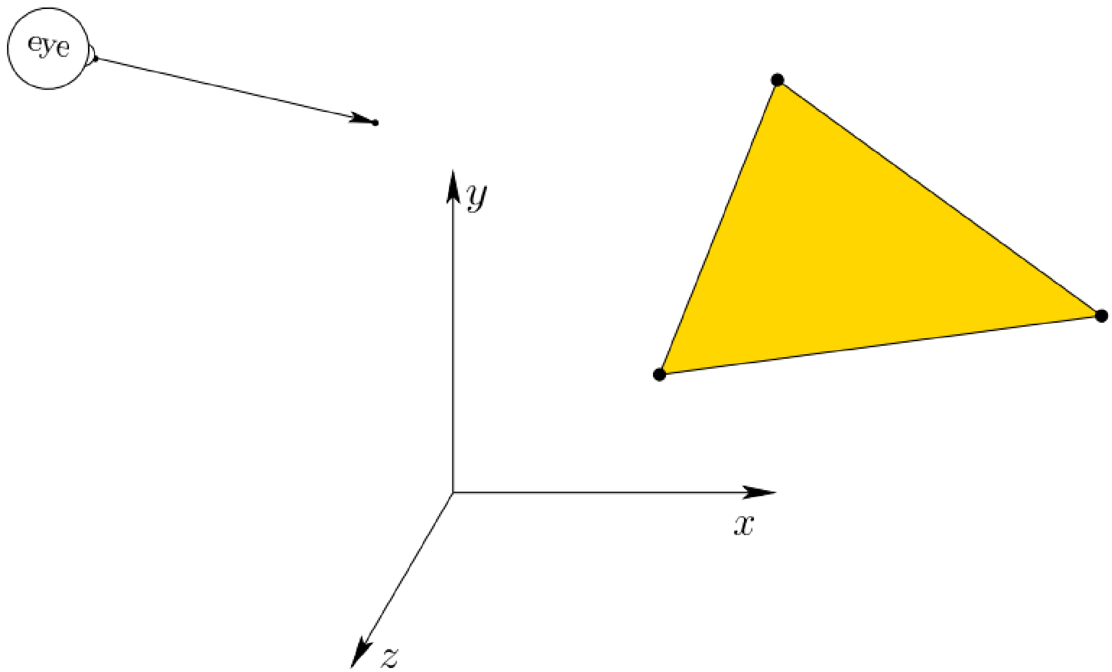
$$T_{rb} = \begin{bmatrix} & R & & x_t \\ & & & y_t \\ & & & z_t \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} & R^T & & -x_t \\ & & & -y_t \\ & & & -z_t \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} & R^T & & 0 \\ & & & 0 \\ & & & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & -x_t \\ 0 & 1 & 0 & -y_t \\ 0 & 0 & 1 & -z_t \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

S. Lavalle

Eye Transform (T_{eye})



Eye position & orientation:

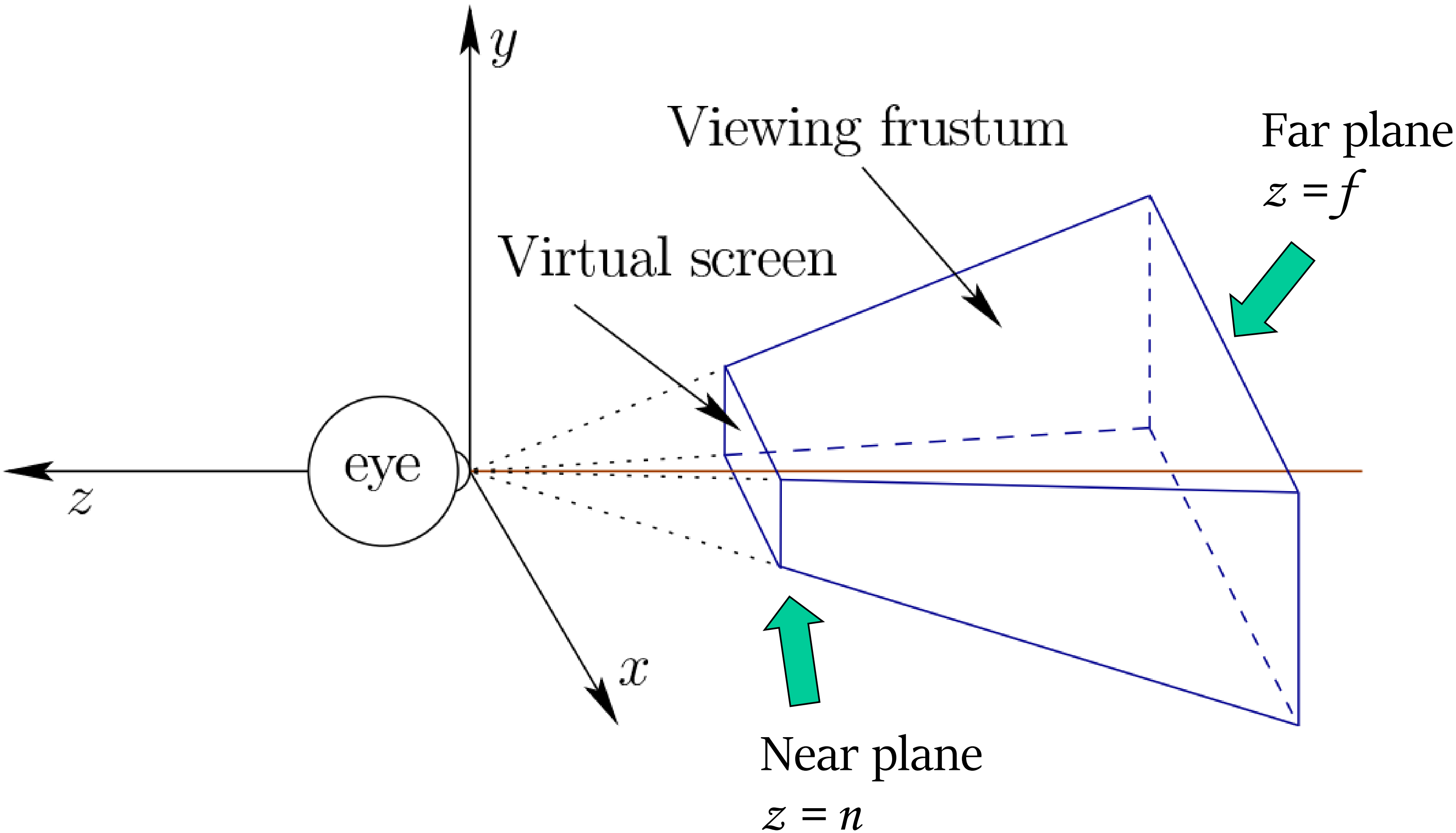
$$e = (e_1, e_2, e_3)$$

$$R_{eye} = \begin{bmatrix} \hat{x}_1 & \hat{y}_1 & \hat{z}_1 \\ \hat{x}_2 & \hat{y}_2 & \hat{z}_2 \\ \hat{x}_3 & \hat{y}_3 & \hat{z}_3 \end{bmatrix}$$

$$T_{eye} = \begin{bmatrix} \hat{x}_1 & \hat{x}_2 & \hat{x}_3 & 0 \\ \hat{y}_1 & \hat{y}_2 & \hat{y}_3 & 0 \\ \hat{z}_1 & \hat{z}_2 & \hat{z}_3 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & -e_1 \\ 0 & 1 & 0 & -e_2 \\ 0 & 0 & 1 & -e_3 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

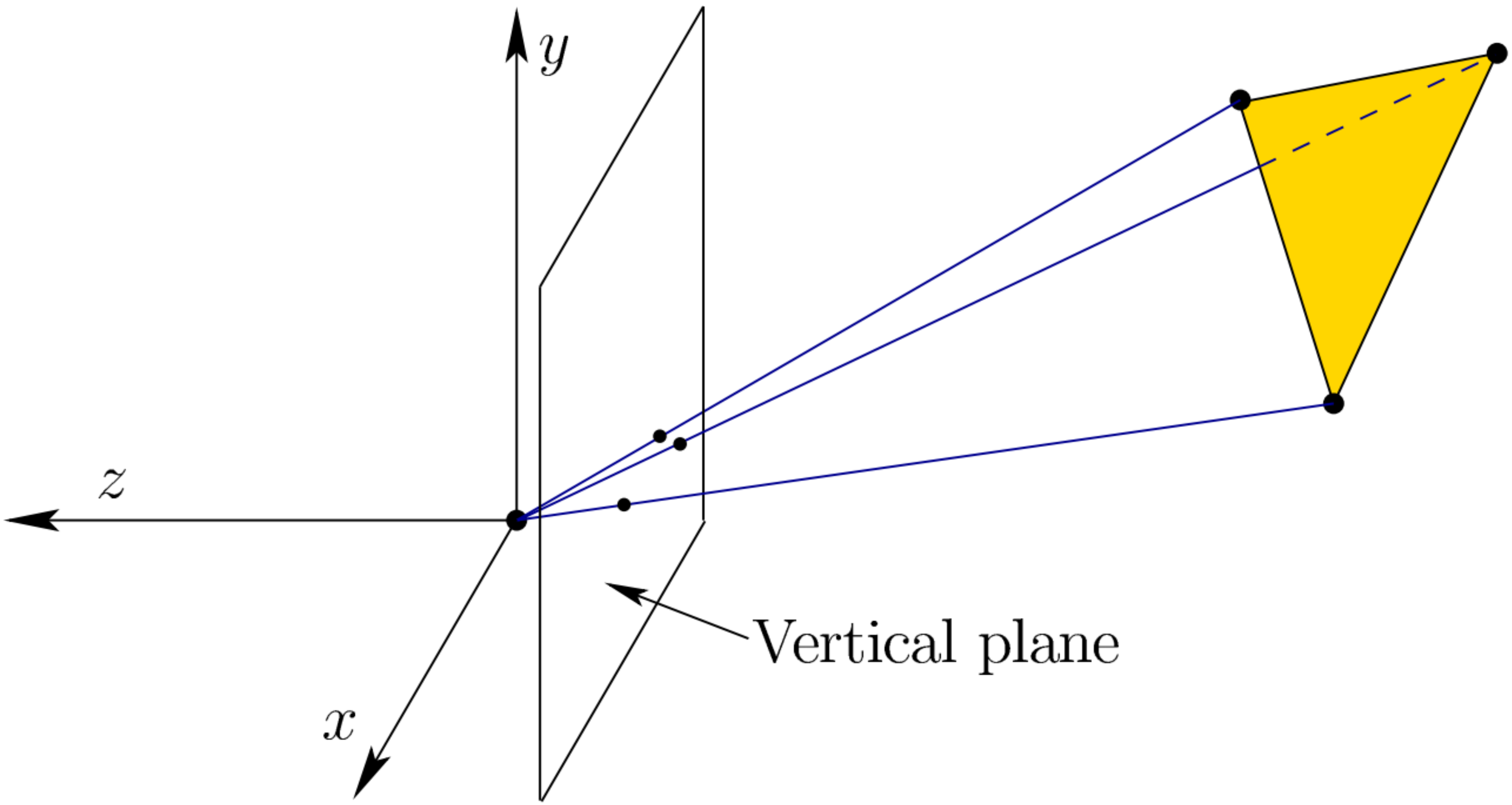
S. Lavalle

Canonical View Transform (T_{can})



S. Lavalle

Perspective Projection



Perspective Projection

Consider this homogeneous transformation matrix:

$$\begin{bmatrix} n & 0 & 0 & 0 \\ 0 & n & 0 & 0 \\ 0 & 0 & n & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} nx \\ ny \\ nz \\ z \end{bmatrix}$$

The resulting 3D coordinates after normalizing:

$$(nx/z, ny/z, n)$$

The first two coordinates describe the projected point, the 3rd is the screen location.

Perspective Projection

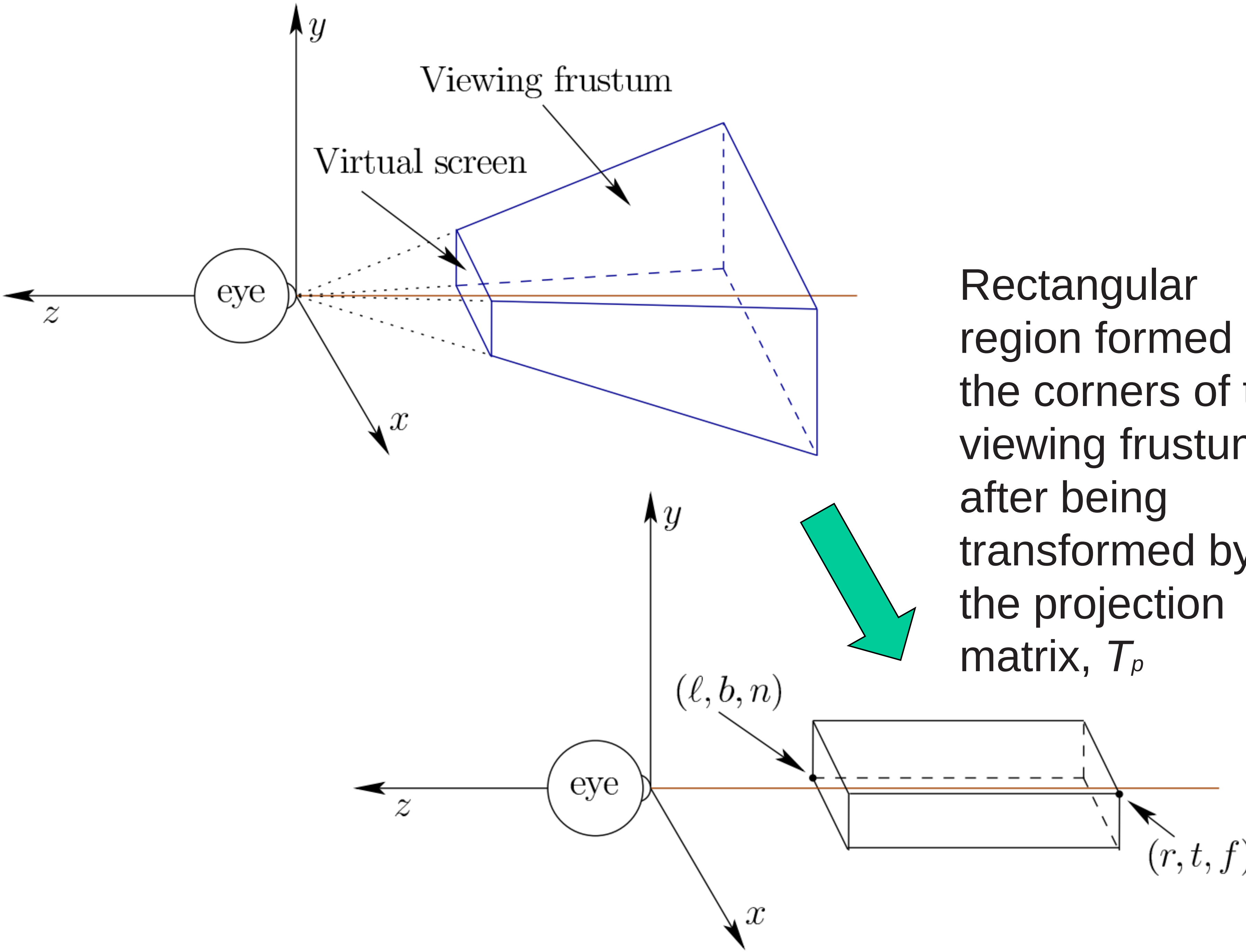
Modification to keep track of "depth" for ordering:

$$\begin{bmatrix} n & 0 & 0 & 0 \\ 0 & n & 0 & 0 \\ 0 & 0 & n & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \Rightarrow T_p = \begin{bmatrix} n & 0 & 0 & 0 \\ 0 & n & 0 & 0 \\ 0 & 0 & n+f & -fn \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

The 3rd coordinate changes to:

$$(n+f)z - fn$$

After normalizing (dividing with z), the depth is preserved for near plane ($z = n$) and far plane ($z = f$). For other values, depth is distorted but the ordering is preserved.



Rectangular region formed by the corners of the viewing frustum, after being transformed by the projection matrix, T_p

Next Step: Translation & Scaling

Translate and rescale the box so that it is centered at the origin and the coordinates of its corners are (±1, ±1, ±1):

$$T_{st} = \begin{bmatrix} \frac{2}{r-\ell} & 0 & 0 & -\frac{r+\ell}{r-\ell} \\ 0 & \frac{2}{t-b} & 0 & -\frac{t+b}{t-b} \\ 0 & 0 & \frac{2}{n-f} & -\frac{n+f}{n-f} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

If the frustum is centered in the xy plane, then the first two components of the last column become 0 .

Canonical view transform becomes: $T_{can} = T_{st} T_p$

Finally: Viewport Transform (T_{vp})

Bring the projected points to the coordinates used to index pixels on the physical display screen:

$$T_{vp} = \begin{bmatrix} \frac{m}{2} & 0 & 0 & \frac{m-1}{2} \\ 0 & \frac{n}{2} & 0 & \frac{n-1}{2} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

m: number of horizontal pixels
n: number of vertical pixels

For example, $n = 1080$ and $m = 1920$ for a 1080p display.

One More Thing: “Cyclopean” to Stereo Views

Assuming interpupillary Distance (IPD) = *t*

$$T_{left} = \begin{bmatrix} 1 & 0 & 0 & \frac{t}{2} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$T = T_{vp} T_{can} T_{eye} T_{rb}$$

$$T = T_{vp} T_{can} T_{left} T_{eye} T_{rb}$$

Similarly, $T_{right} = \begin{bmatrix} 1 & 0 & 0 & -\frac{t}{2} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

S. Lavalle