Origins of structure in perception

Sarah Schwettmann
Vision in Art and Neuroscience
2023

Disappearance of structure in perception

STABILIZED IMAGES typically fade as in the illustrations on this and the following two pages. The parts of a profile drawing that stay visible are invariably specific features or groups of features, such as the front of the face or the top of the head.

MEANINGLESS CURLICUES first come and go in random sequence. But after a while small groups of curlicues organized in recognizable patterns start to behave as units. This suggests that they have themselves become meaningful perceptual elements.

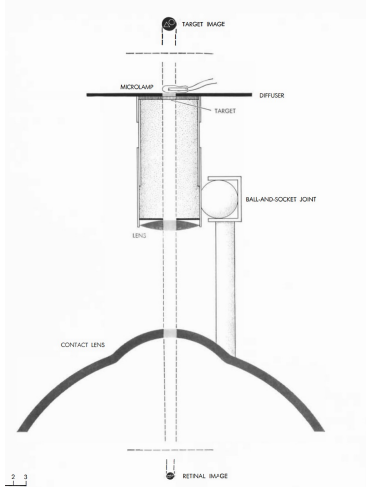
MONOGRAM formed of the letters *H* and *B* also seems to illustrate the importance of elements that are meaningful because of past experience. When the monogram breaks up it is the recognizable letters and numbers within it that come successively into view.

WORDS containing other words behave in much the same manner as the monogram. Here, for example, the subject sees new words made up of letters and parts of letters in the original. He is far less likely to report seeing meaningless groups of letters such as *EER*.

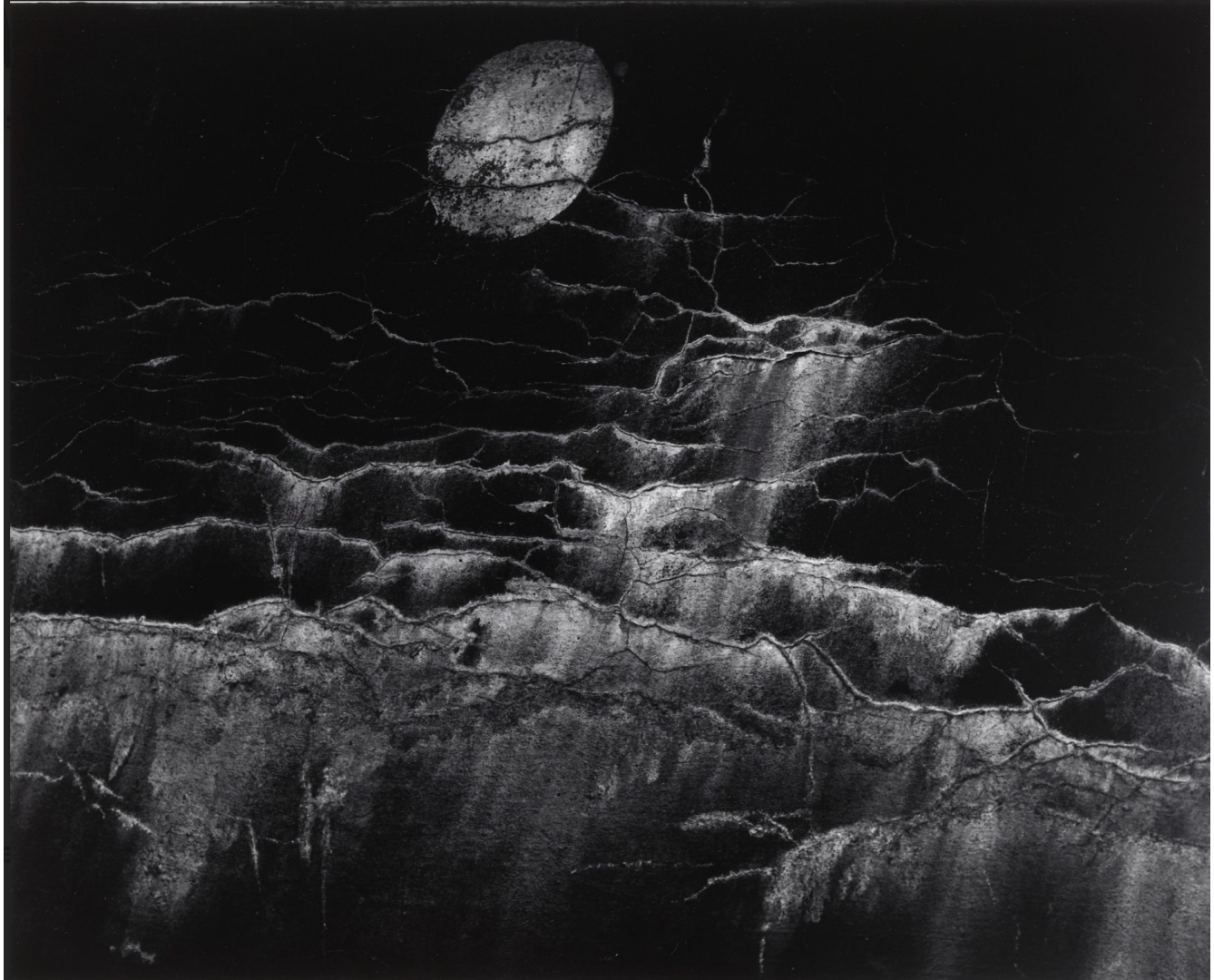
Stabilized Images on the Retina

When the involuntary movements of an image across the retina are prevented, the image fades and reappears in a manner that provides new information on two major theories of perception

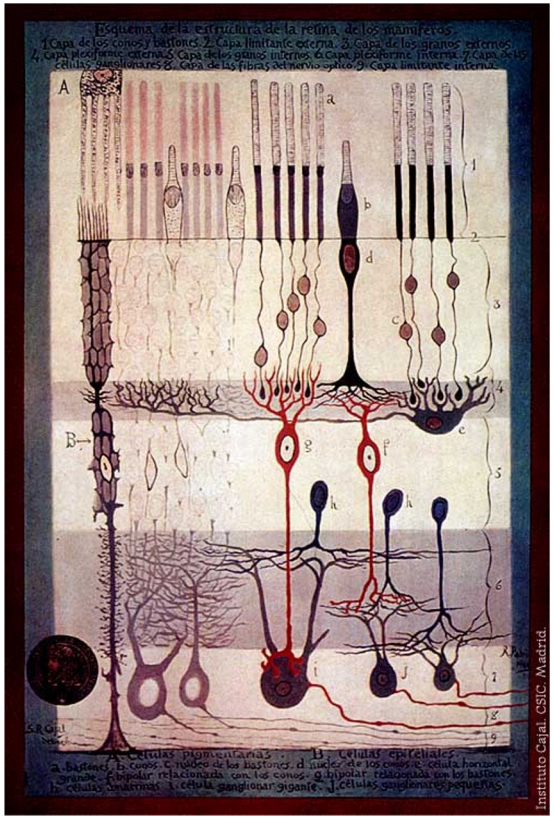
by Roy M. Pritchard

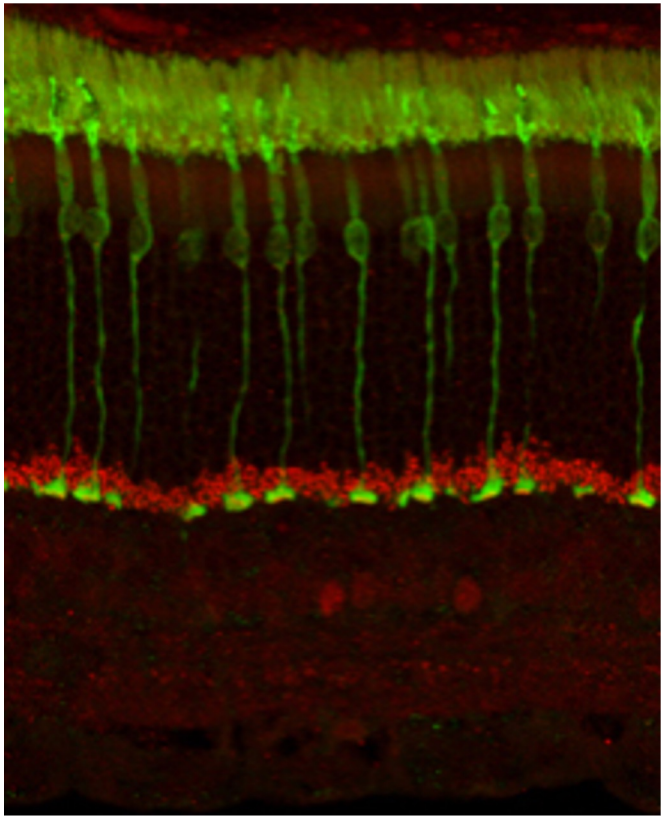


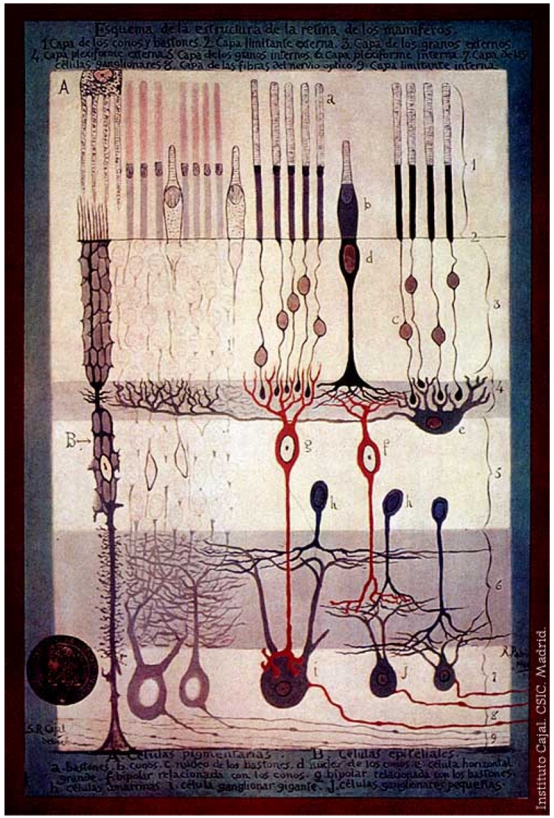
Minor White *Windowsill Daydreaming*
July 1958

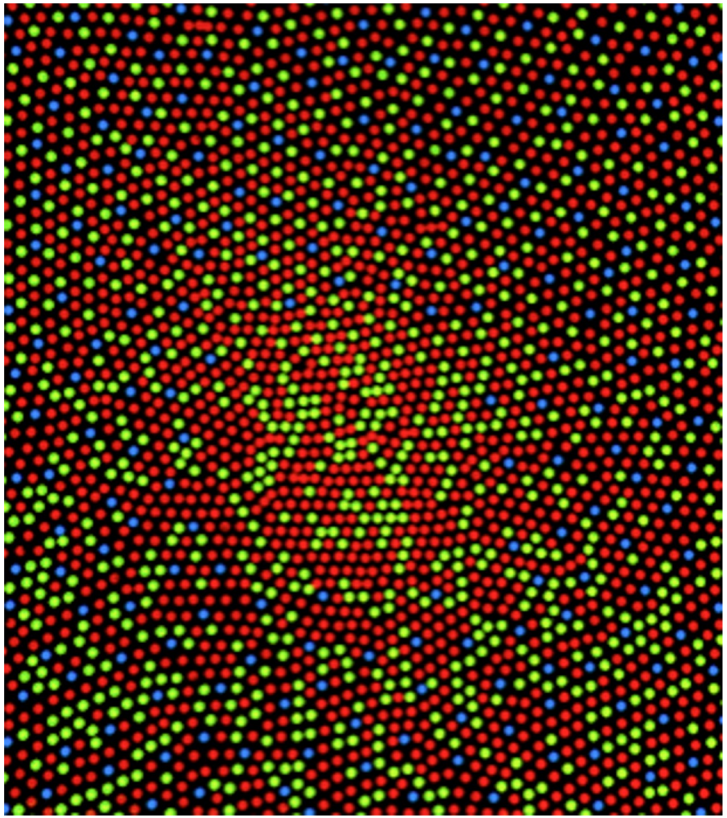



THE ULTIMATE ADHESIVE
FEVICOL


Esquema de la estructura de la retina de los mamíferos.
1.Capa de los conos y bastones. 2. Capa limitante externa. 3. Capa de los granos externos 4. Capa plexiforme externa 5. Capa de los granos internos 6. Capa plexiforme interna. 7. Capa de las células ganglionares 8. Capa de las fibras del nervio óptico 9. Capa limitante interna.
A
B
a
b
c
d
e
f
g
h
i
j
1
2
3
4
5
6
7
8
9
S.R.Cajal
A. Células pigmentarias. B. Células epiteliales.
a. Bastones. b. Conos. c. Núcleo de los bastones. d. Núcleo de los conos. e. Célula horizontal. Grande. f. Bipolar relacionada con los conos. g. Bipolar relacionada con los bastones. h. Células centrales. i. Célula ganglionar gigante. j. Células ganglionares pequeñas.
Instituto Cajal. CSIC. Madrid.



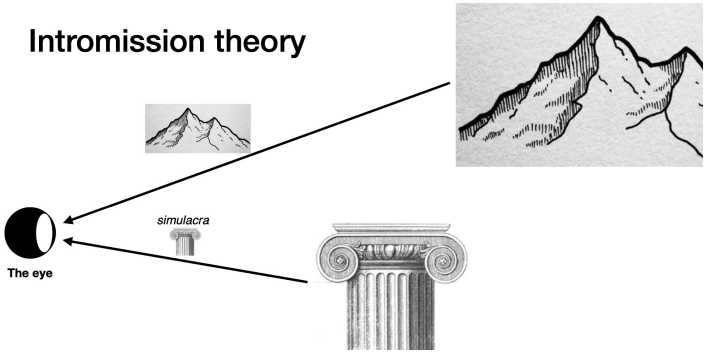

Esquema de la estructura de la retina de los mamíferos.
1.Capa de los conos y bastones. 2. Capa limitante externa. 3. Capa de los granos externos
4. Capa plexiforme externa 5. Capa de los granos internos. 6. Capa plexiforme interna. 7. Capa de las
células ganglionares 8. Capa de las fibras del nervio óptico 9. Capa limitante interna.
A
B
a
b
c
d
e
f
g
h
i
j
1
2
3
4
5
6
7
8
9
A. Células pigmentarias. B. Células epiteliales.
a. bastones. b. conos. c. cuerpo de los bastones. d. núcleo de los conos. e. célula horizontal.
f. grande célula bipolar, relacionada con los conos. g. bipolar, relacionada con los bastones.
h. células centrales. i. célula ganglionar gigante. j. células ganglionares pequeñas.
S.R.Cajal
Instituto Cajal. CSIC. Madrid.



How do the eyes receive information about the world?

425 BCE: Intromission theories

- Objects emit copies of themselves (eidola or simulacra) that enter the eyes
- Advanced by Epicurus, Lucretius, Demokritos, among others
- How could objects send copies to multiple observers?
  How did the copies not interfere with one another?
  How could copies of large objects *fit into the eye*?



~400 BCE: Extramission theories

- Empedocles, Plato, Euclid
- Empedocles: all things are composed of 4 elements: fire, earth, air, water. The eye contains the 4 elements and **fire is responsible for creating rays that emanate from the eye and touch the world** (this fire is is why cats' eyes glow!)
- Later, Aristotle:
  -stars too far away for rays from the eyes to reach them!
  -some objects are light sources and others reflect rays that reach the eyes
  -**water** is the element of perception as vision needed a transparent element!
  (Aristotle, *On Sense and the Sensible,* 350 BCE*)*

Light travels along straight lines – coming from where?

325 BCE: Euclid, a first mathematical theory of vision
(in *Optics*)

- Rays emitted *by the eye* traveled along a straight line and formed a cone reaching the scene
- Work set basis for **perspective**

Nothing that is seen is seen at once in its entirety. (Fig. 1). For let the thing seen be AD, and let the eye be B, from which let the rays of vision fall, BA, BG, BK, and BD. So, since the rays of vision, as they fall, diverge from one another, they could not fall in continuous line upon AD; so that there would be spaces also in AD upon which the rays of vision would not fall. So AD will not be seen in its entirety at the same time. But it seems to be seen all at once because the rays of vision shift rapidly.

Objects located nearby are seen more clearly than objects of equal size located at a distance. (Fig. 2.)

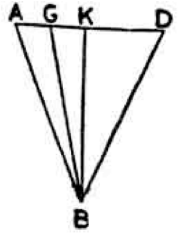


FIG. 1.

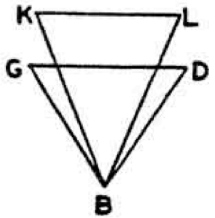


FIG. 2.

From Euclid's "Optics" (325 BC)
https://philomatica.org/wp-content/uploads/2013/01/Optics-of-Euclid.pdf

Light travels along straight lines – coming from where?

325 BCE: Euclid, a first mathematical theory of vision (in *Optics*)

- Rays emitted *by the eye* traveled along a straight line and formed a cone reaching the scene
- Work set basis for **perspective**

965-1040 CE: Ibn al-Haytham (“Alhazen”)

- *Book of Optics* (1028-1038 CE): light rays bounce off objects in all directions and become visible when they reach the eye perpendicularly

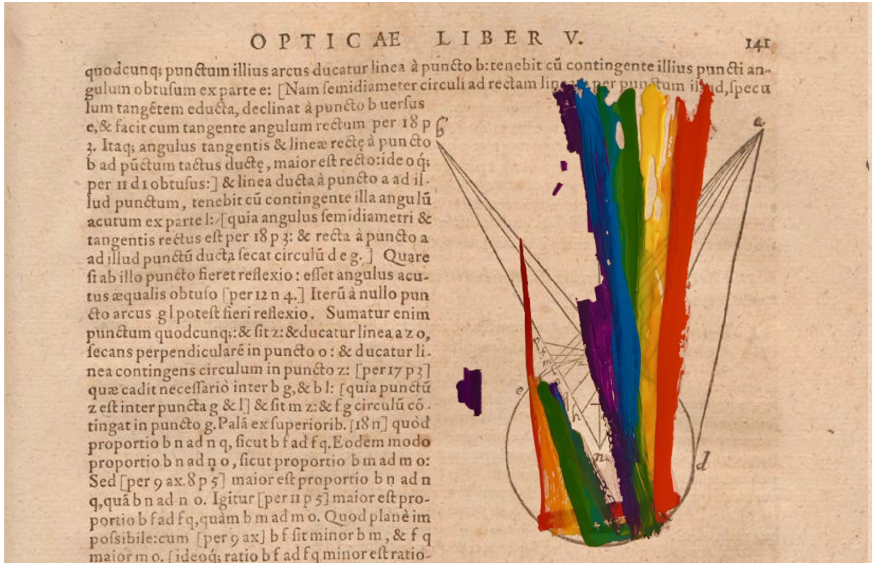
Light travels along straight lines – coming from where?

325 BCE: Euclid, a first mathematical theory of vision (in *Optics*)

- Rays emitted *by the eye* traveled along a straight line and formed a cone reaching the scene
- Work set basis for **perspective**

965-1040 CE: Ibn al-Haytham (“Alhazen”)

- *Book of Optics* (1028-1038 CE): light rays bounce off objects in all directions and become visible when they reach the eye perpendicularly



From “Total Internal Reflection” – VAN 2020, https://vision.mit.edu/tir

Helmholtz: Perception as Inference

1821-1894: *Hermann von Helmholtz*

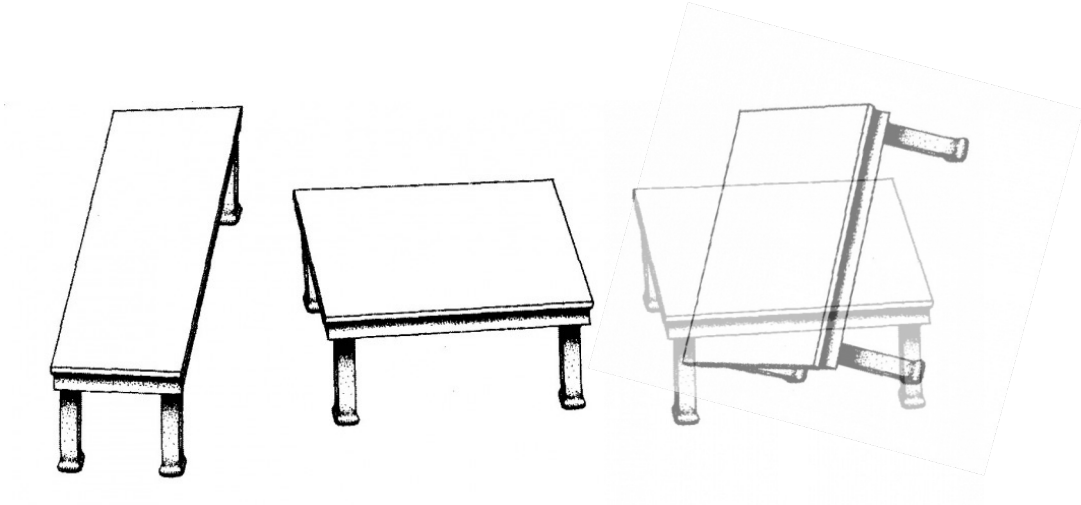
> The general rule determining the ideas of vision that are formed whenever an impression is made on the eye, is that *such objects are always imagined as being present in the field of vision as would have to be there in order to produce the same impression on the nervous mechanism, the eyes being used under ordinary normal conditions.*

Treatise on Physiological Optics (1866)

- The mind makes perceptions out of sensations finding representations of the object most likely to explain the sensory input.
- Related to Bayesian methods for computer vision inference.

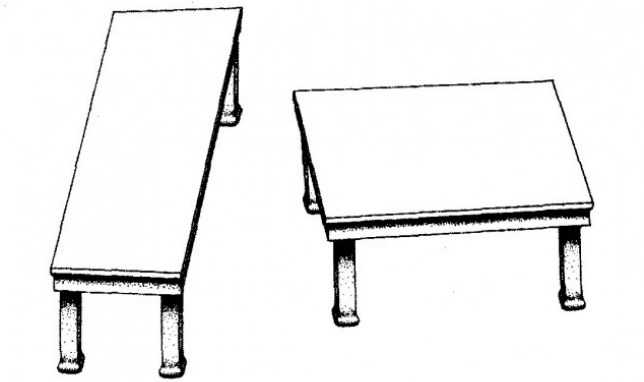
The human visual system recovers scenes that are the *cause* of images

The human visual system recovers scenes that are the *cause* of images

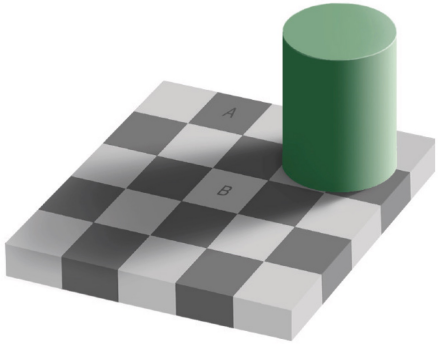


Roger Shephard's Turning Tables (1990)

The human visual system recovers scenes that are the *cause* of images



Roger Shephard's Turning Tables (1990)



Adelson's Checkershadow Illusion (1995)

Perception as inference and explanation, not as measurement

The human visual system recovers scenes that are the *cause* of images

Hoarfrost at Ennery, Camille Pissarro. Oil on canvas, 1873

The human visual system recovers scenes that are the *cause* of images

Helmholtz 1871 lecture

"We must look upon artists as persons whose observation of sensuous impressions is particularly vivid and accurate, and whose memory for these images is particularly true … The study of works of art will throw great light on the question as to which elements and relations of our visual impressions are most predominant in determining our conception of what is seen." (Helmholtz, 1995) (p.280), quoted in (Hyman, 2010).

Hoarfrost at Ennery, Camille Pissarro. Oil on canvas, 1873

These scrape marks are sufficient to convey the appearance of a heavily frosted field.

The human visual system recovers scenes that are the *cause* of images – how?

Origins of Structure in Perception: this will take us a semester to answer!

- Low level vision
- Motion and depth
- Color
- Recognition
- Association

And in parallel:

"We must look upon artists as persons whose observation of sensuous impressions is particularly vivid and accurate, and whose memory for these images is particularly true … The study of works of art will throw great light on the question as to which elements and relations of our visual impressions are most predominant in determining our conception of what is seen." (Helmholtz, 1995) (p.280), quoted in (Hyman, 2010).

Perception as an inverse problem

Pizlo 2000

“perception is about inferring the properties of the distal stimulus X given the proximal stimulus Y”

$$Y = AX \quad \Rightarrow \quad X = A^{-1}Y$$

where

- Y is the proximal stimulus (e.g. retinal image)
- X is the distal stimulus (e.g. 3D object)
- A is a perspective mapping (linear transformation)

ill-posed, determining a unique A^{-1} is difficult

many different things could have created the same retinal image (e.g. of a cube)

Less clear why binocular reconstruction of a 3D scene is ill-posed
(e.g. Gibson: observer knows position of one eye relative to the other, reconstruction is well-posed and well-conditioned)

Problem: perceptual and motor noise

- Rays never actually intersect
- One could approximate the intersection, but the approximation is very unstable (small variations in the noise cause large variations in the solution)

Perception as an inverse problem

Pizlo 2000

“perception is about inferring the properties of the distal stimulus X given the proximal stimulus Y”

$$Y = AX \quad \Rightarrow \quad X = A^{-1}Y$$

where

Y is the proximal stimulus (e.g. retinal image)

X is the distal stimulus (e.g. 3D object)

A is a perspective mapping (linear transformation)

ill-posed, determining a unique A^{-1} is difficult

many different things could have created the same retinal image (e.g. of a cube)

to solve: $X = A^{-1}Y$ ⟹ regularized linear regression:

$$E_1 = \left\| AX - Y_X \right\|^2 + \lambda \left\| P_X \right\|^2$$

Bayes’ rule:

$$p(X|Y_X) = \frac{p(Y_X|X)p(X)}{p(Y_X)}$$

$$-\log\ p(X|Y_X) = -\log\ p(Y_X|X) - \log\ p(X)$$

Predictive Coding

"the brain is in the game of optimizing neuronal dynamics and connectivity to maximize the evidence for its model of the world" (Friston 2018, "Does Predictive Coding Have a Future?")

- neuronal representations in higher/deeper levels of hierarchies generate predictions about representations in lower levels
- descending predictions are compared with lower-level representations to form a prediction error, which is passed back up the hierarchy to update higher-level representations
- The recurrent exchange of signals between adjacent levels encodes a *generative model* – generating predictions of sensations that can be compared with actual sensory samples

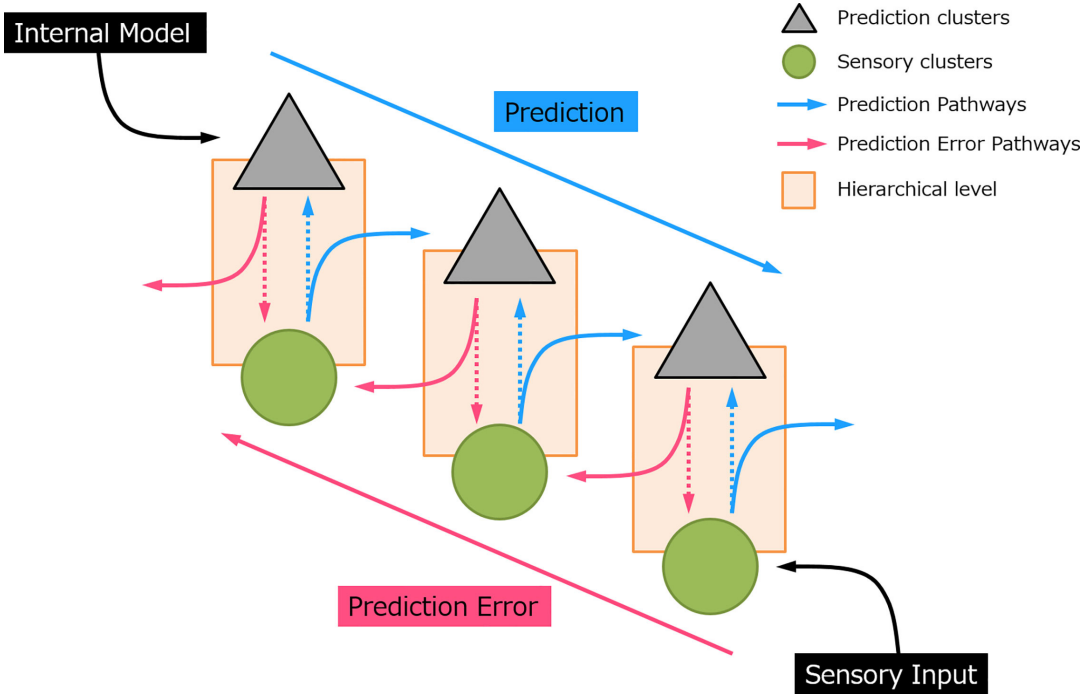
Published: January 1999

Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects

Rajesh P. N. Rao & Dana H. Ballard

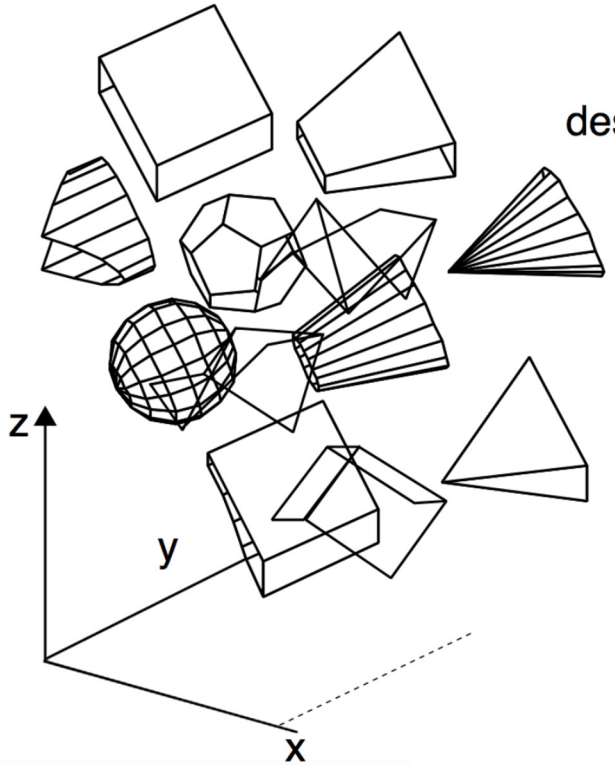
Nature Neuroscience **2**, 79–87 (1999) | Cite this article

41k Accesses | **2756** Citations | **112** Altmetric | Metrics



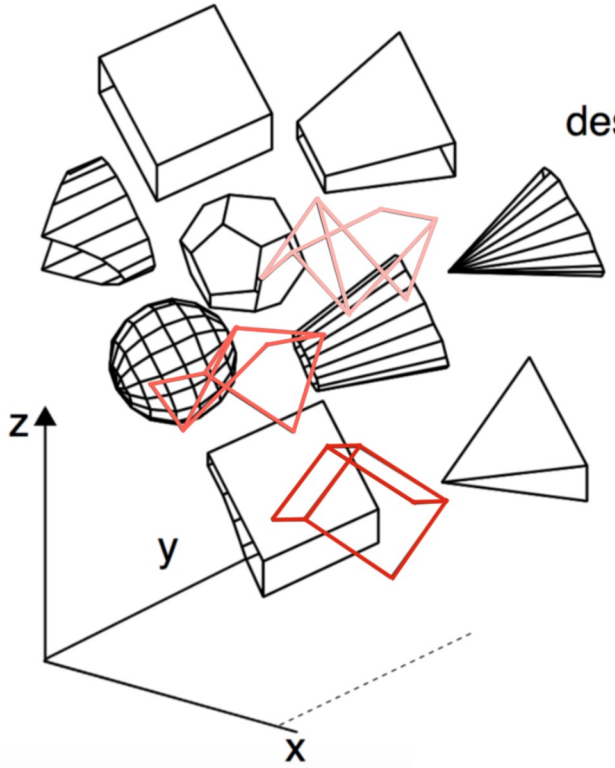
From Fong et al. 2020

Bayesian Models of Object Perception



"It seems reasonable to assume that an effective perceptual system should obtain and use knowledge of the constraints that characterize the distal stimuli to 'make up' for the information that has been lost in the transformation from the distal to proximal stimulus." (Pizlo 2000)

adapted from Kersten & Yuille (2003)

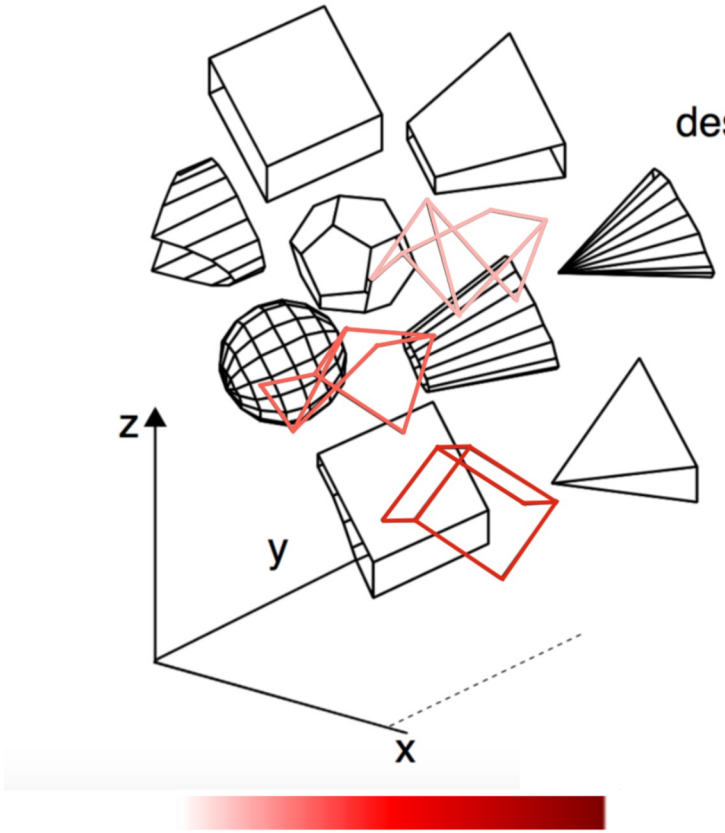


unlikely likely

prior probability

to solve: $X = A^{-1}Y$ constraints!

adapted from Kersten & Yuille (2003)



object descriptions, S

unlikely likely

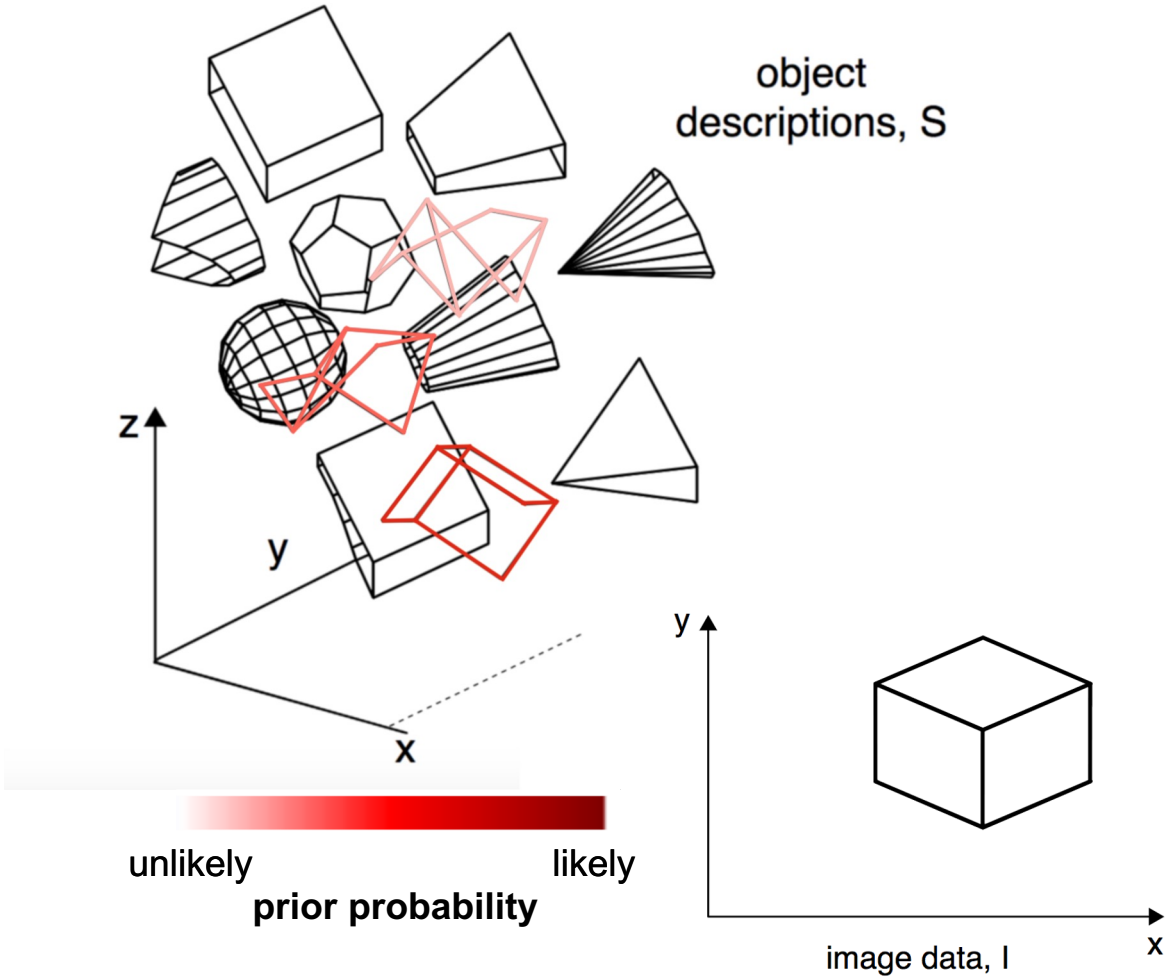
prior probability

image data, I

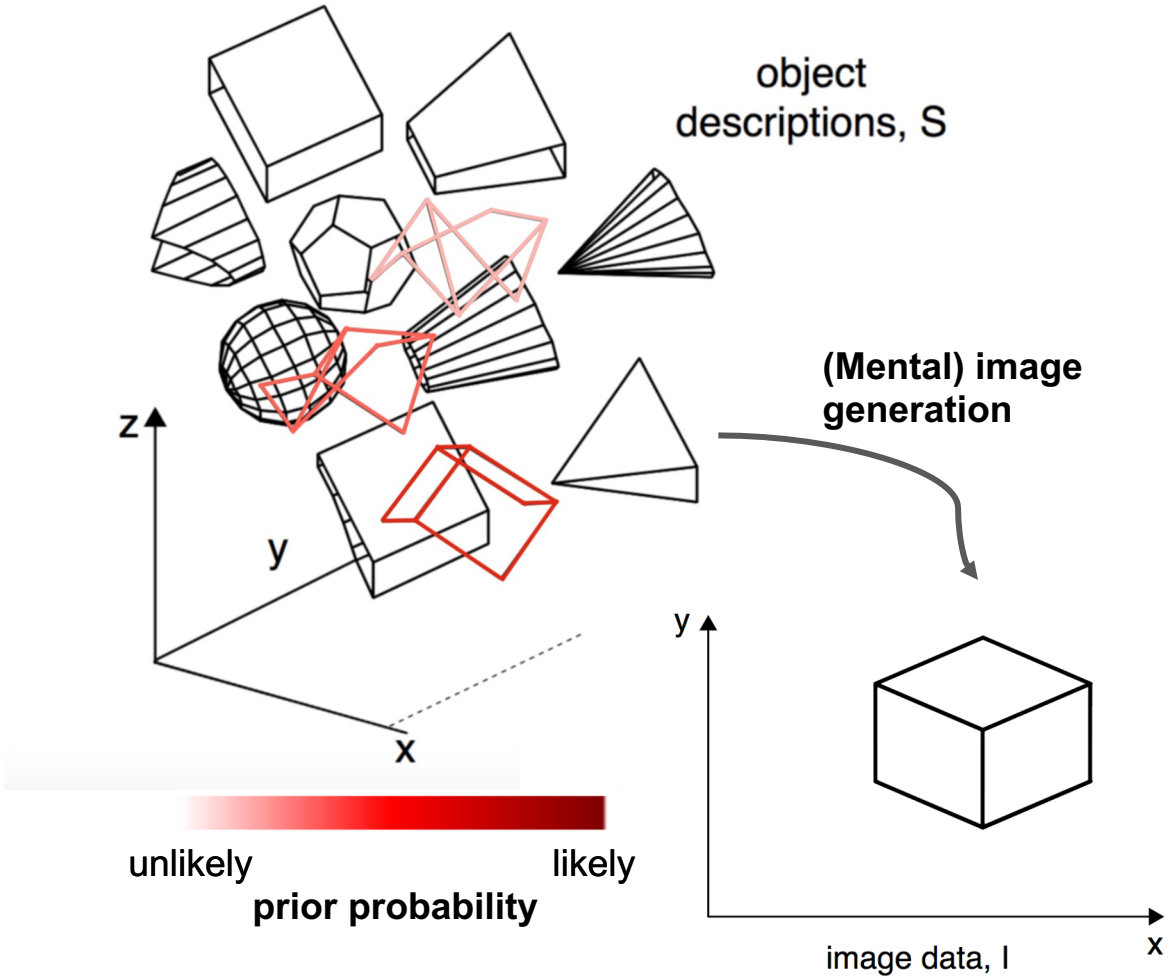
telling the robot what they are seeing. If you could see the world through a robot's eyes, it would look not like a movie picture decorated with crosshairs but something like this:

225	221	216	219	219	214	207	218	219	220	207	155	136	135
213	206	213	223	208	217	223	221	223	216	195	156	141	130
206	217	210	216	224	223	228	230	234	216	207	157	136	132
211	213	221	223	220	222	237	216	219	220	176	149	137	132
221	229	218	230	228	214	213	209	198	224	161	140	133	127
220	219	224	220	219	215	215	206	206	221	159	143	133	131
221	215	211	214	220	218	221	212	218	204	148	141	131	130
214	211	211	218	214	220	226	216	223	209	143	141	141	124
211	208	223	213	216	226	231	230	241	199	153	141	136	125
200	224	219	215	217	224	232	241	240	211	150	139	128	132
204	206	208	205	233	241	241	252	242	192	151	141	133	130
200	205	201	216	232	248	255	246	231	210	149	141	132	126
191	194	209	238	245	255	249	235	238	197	146	139	130	132
189	199	200	227	239	237	235	236	247	192	145	142	124	133
198	196	209	211	210	215	236	240	232	177	142	137	135	124
198	203	205	208	211	224	226	240	210	160	139	132	129	130
216	209	214	220	210	231	245	219	169	143	148	129	128	136
211	210	217	218	214	227	244	221	162	140	139	129	133	131
215	210	216	216	209	220	248	200	156	139	131	129	139	128
219	220	211	208	205	209	240	217	154	141	127	130	124	142
229	224	212	214	220	229	234	208	151	145	128	128	142	122
252	224	222	224	233	244	228	213	143	141	135	128	131	129
255	235	230	249	253	240	228	193	147	139	132	128	136	125
250	245	238	245	246	235	235	190	139	136	134	135	126	130
240	238	233	232	235	255	246	168	156	144	129	127	136	134

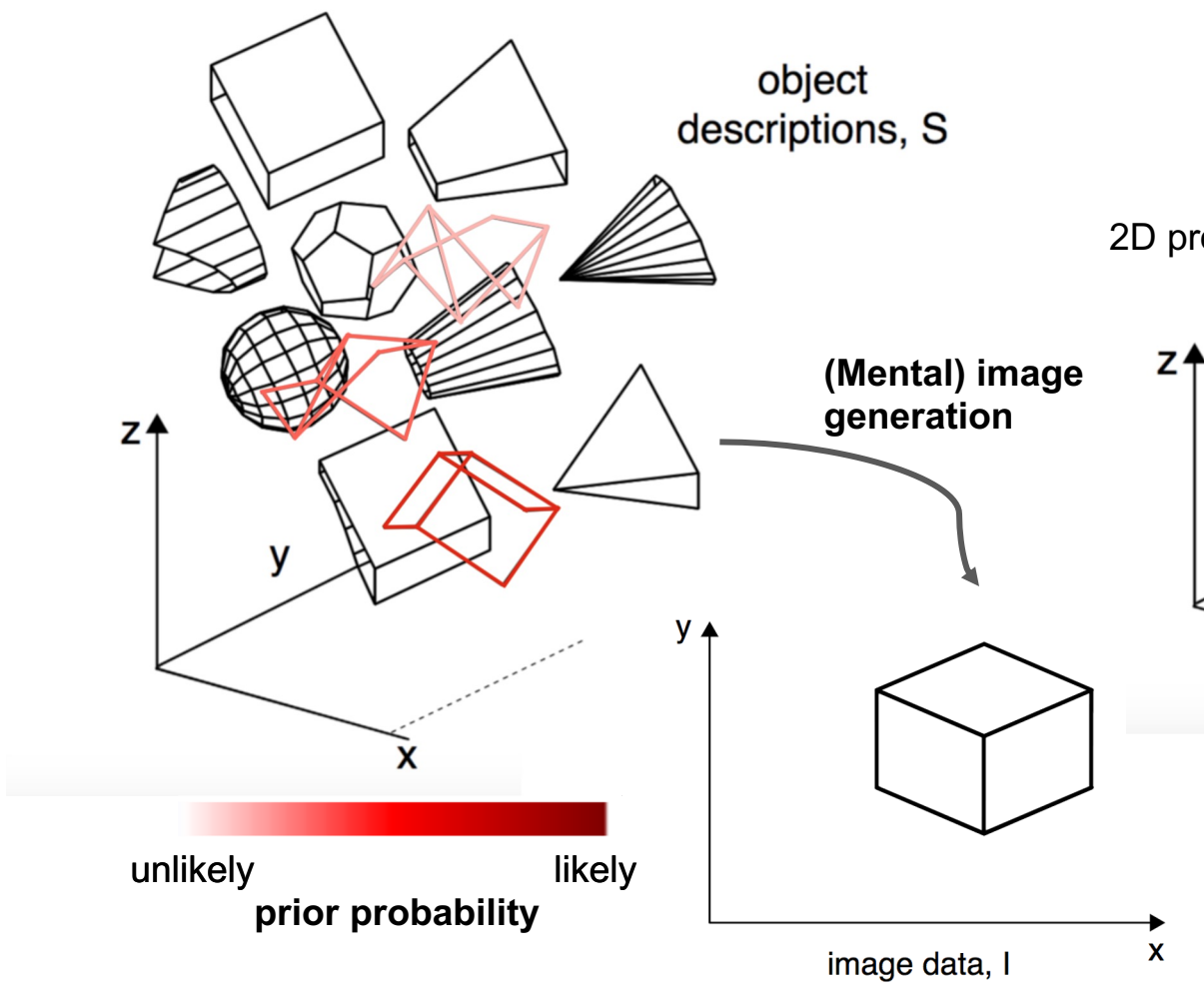
adapted from Kersten & Yuille (2003)



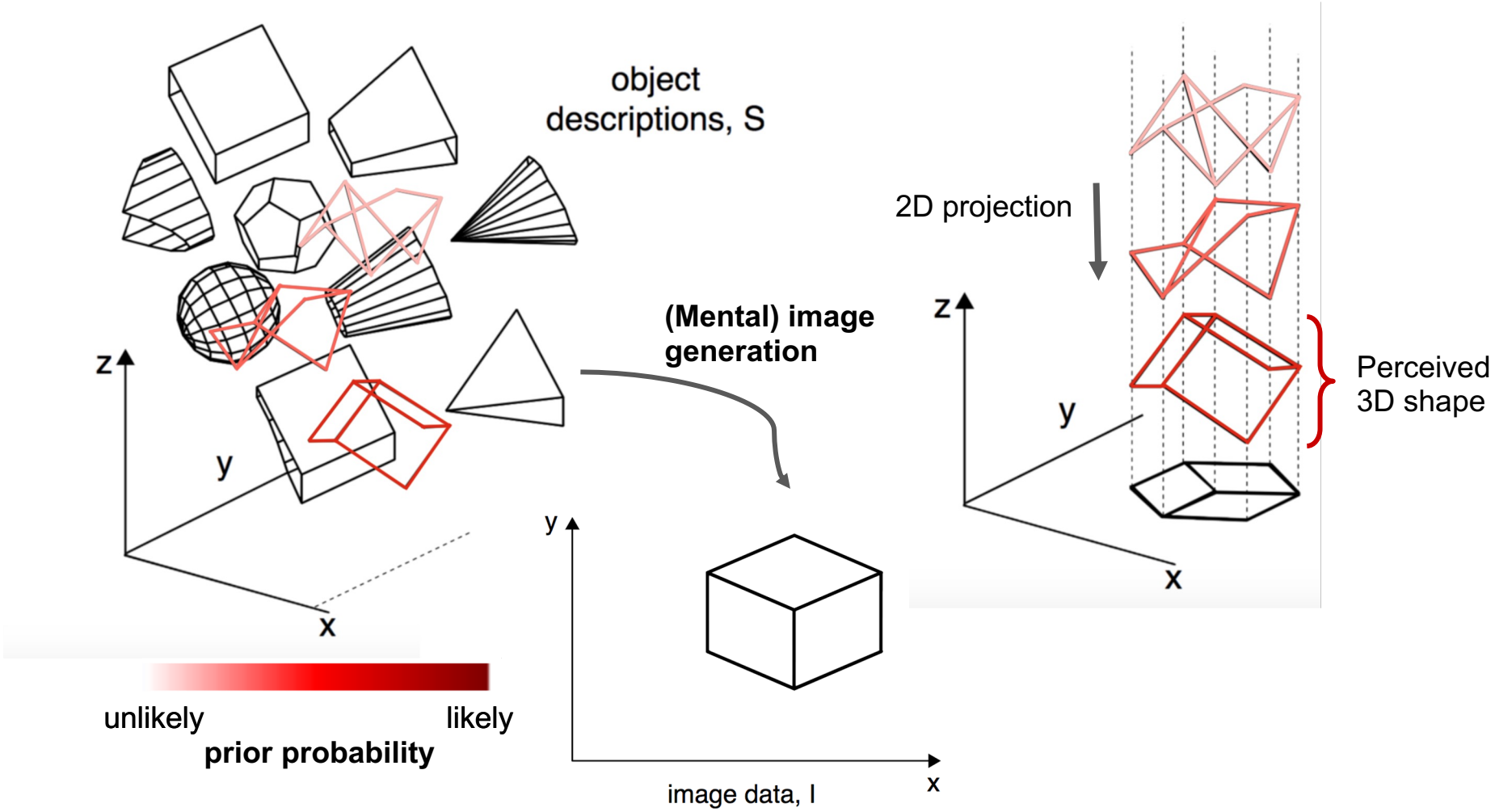
adapted from Kersten & Yuille (2003)



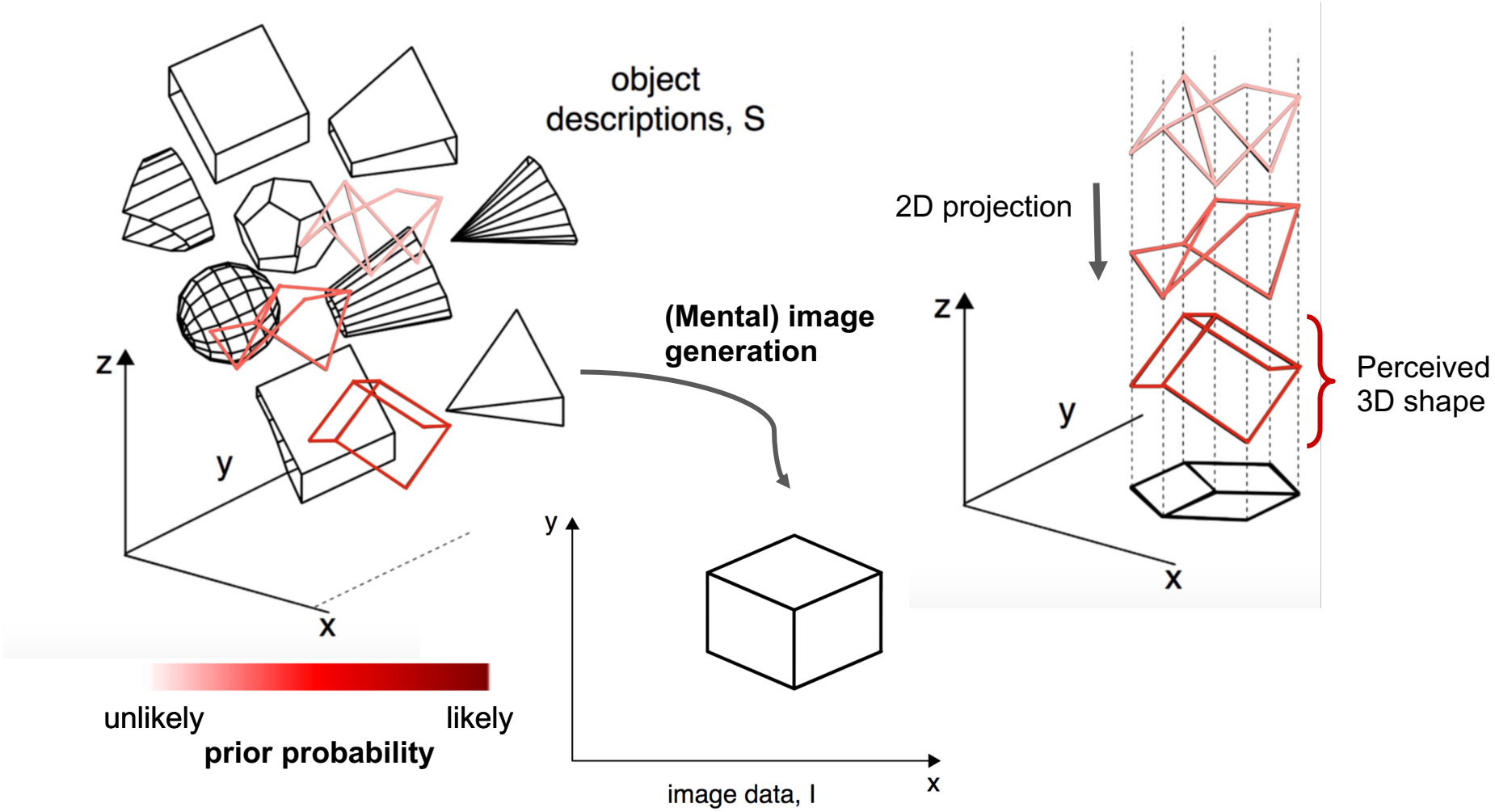
adapted from Kersten & Yuille (2003)



adapted from Kersten & Yuille (2003)



adapted from Kersten & Yuille (2003)

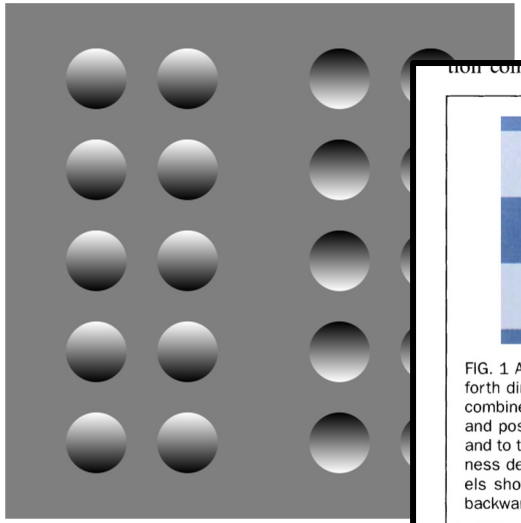


adapted from Kersten & Yuille (2003)

Visual system embodies principles of ecological optics

- Lighting from above, shadows, reflection, ...

“evolutionary internalized regularities”



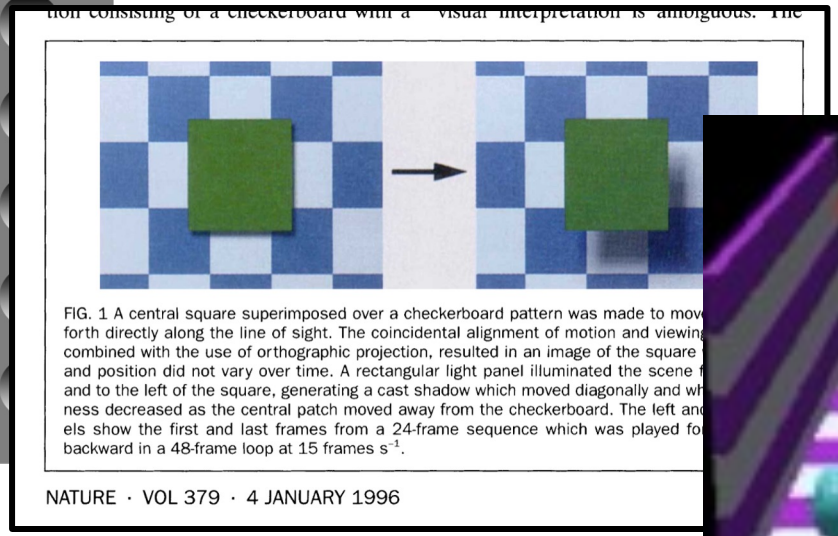
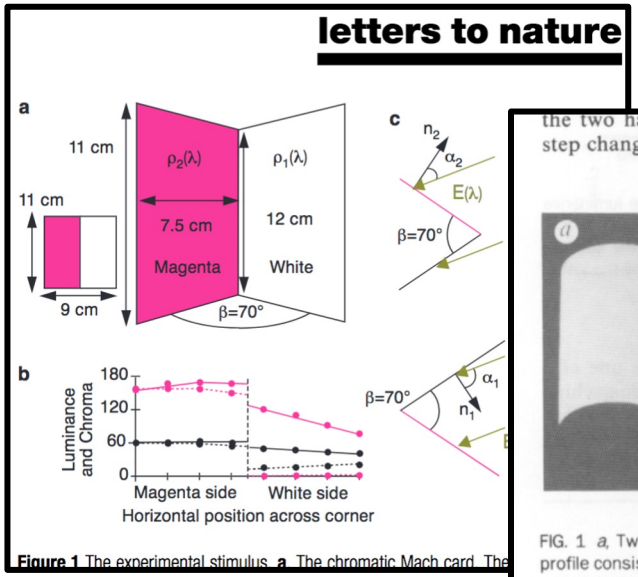


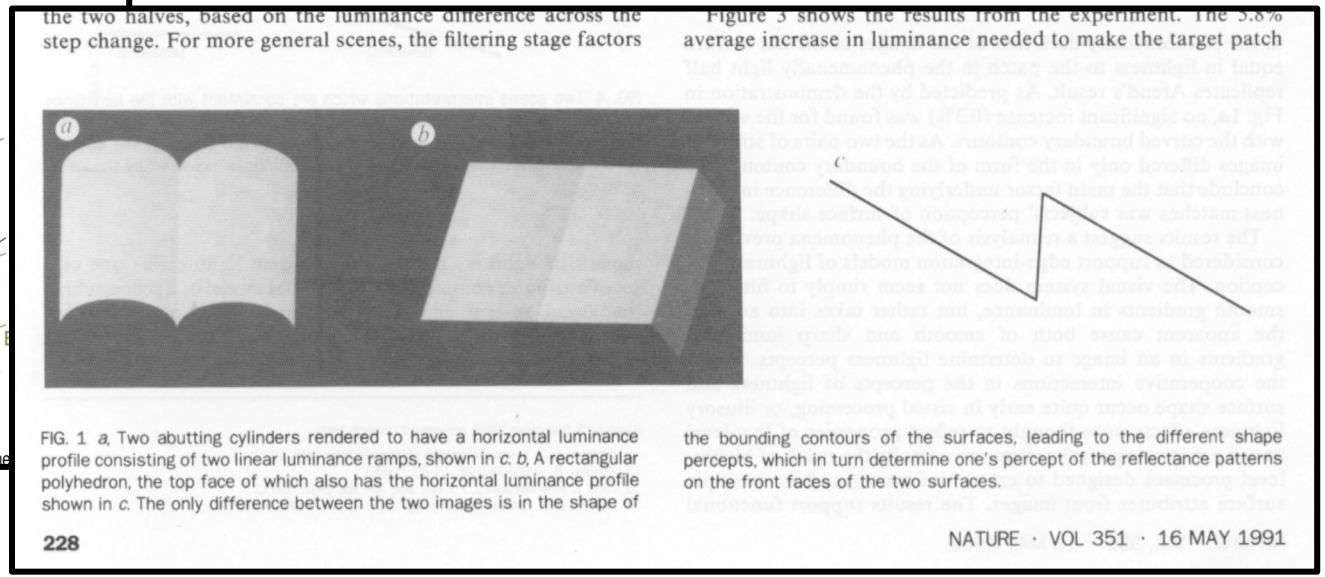
FIG. 1 A central square superimposed over a checkerboard pattern was made to mov
forth directly along the line of sight. The coincidental alignment of motion and viewin
combined with the use of orthographic projection, resulted in an image of the square
and position did not vary over time. A rectangular light panel illuminated the scene
and to the left of the square, generating a cast shadow which moved diagonally and wh
ness decreased as the central patch moved away from the checkerboard. The left and
els show the first and last frames from a 24-frame sequence which was played fo
backward in a 48-frame loop at 15 frames s^{-1}.

NATURE · VOL 379 · 4 JANUARY 1996

Visual system embodies principles of ecological optics

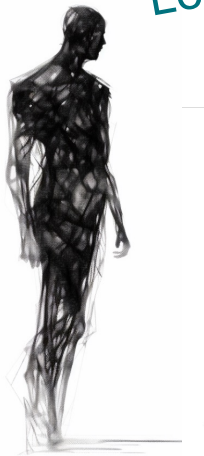
- Interactions of shape and illumination

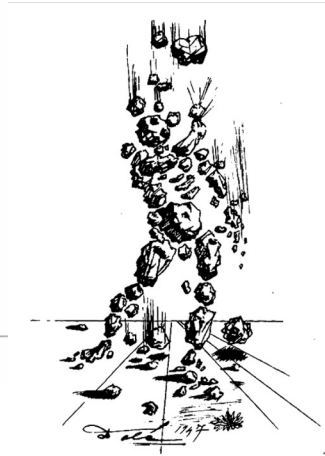




Learned (?) Models in Art

images from Midjourney (2023)





possible renderings of a common shape are infinite but no problem!

40

Koenderink (2014)

clarence coles phillips (1880-1927)

The Beholder's Share

1909-2001: Ernst Gombrich, Austrian art historian

"draws [the beholder] into the magic circle of creation and allows him to experience something of the thrill of 'making' which had once been the privilege of the artist" (Gombrich, 1961) (p. 202).

Clarence Coles Phillips

Camille Pissarro

Paul Cezanne

The Beholder's Share

Perceptual Metamerization:

- Stimuli that differ physically but look the same (from a particular distance and point of fixation)
- Creation requires a model of perceptual processing
- For example, spatial pooling, both in successive ventral stream areas, and with eccentricity, induces an irretrievable loss of information:

Nat Neurosci. Author manuscript; available in PMC 2012 Mar 1.

Published in final edited form as:

Nat Neurosci. 2011 Sep; 14(9): 1195–1201.

Published online 2011 Aug 14. doi: 10.1038/nn.2889

Metamers of the ventral stream

Jeremy Freeman[1] and Eero P. Simoncelli[1,2,3]

The Beholder's Share

1909-2001: Ernst Gombrich, Austrian art historian

> "draws [the beholder] into the magic circle of creation and allows him to experience something of the thrill of 'making' which had once been the privilege of the artist" (Gombrich, 1961) (p. 202).

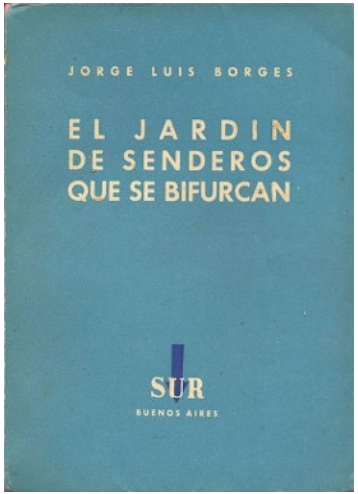
Koenderink 2012:

QUESTION: Is perception like a bundle of possible visual worlds (then the observer apparently abstains from a final – precognitive – decision) or is it like a single visual world (the observer really "sticks the neck out") ?

In other words, do observers resolve ambiguities when not required for an action or decision?

Minor White *Windowsill Daydreaming*
July 1958

QUESTION: Is perception like a bundle of possible visual worlds (then the observer apparently abstains from a final – precognitive – decision) or is it like a single visual world (the observer really "sticks the neck out") ?



I think that in a great many cases perceptions are more of the many visual worlds variety than like the single guess. You don't notice this in the laboratory because most psychophysical methods *force a unique response*. They simulate the decisions–in–action of daily life. That you don't notice the essential ambiguity of perception in real life is most likely due to the fact that you don't need to take decisions on issues on which no actions will be taken anyway. That the many visual worlds option is indeed likely is suggested by the fact that a change of psychophysical method or task often leads to distinctly different results. This is not to say that observers actually entertain many visual worlds interpretations explicitly, but merely that they don't necessarily resolve ambiguities when this is not specifically required for some action or decision.

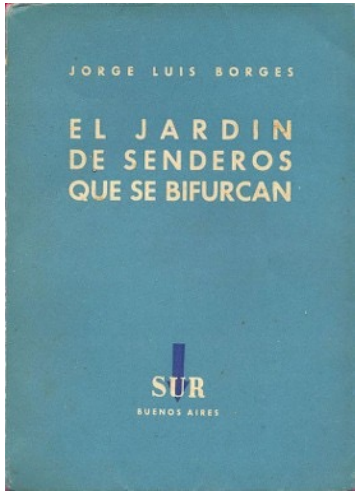
Multiple Worlds (2012)

Consciousness: facing the consequences of decisions-in-action? (Schrodinger)

Bev Doolittle, *Pintos* (2000)

I think that in a great many cases perceptions are more of the many visual worlds variety than like the single guess. You don't notice this in the laboratory because most psychophysical methods *force a unique response*. They simulate the decisions–in–action of daily life. That you don't notice the essential ambiguity of perception in real life is most likely due to the fact that you don't need to take decisions on issues on which no actions will be taken anyway. That the many visual worlds option is indeed likely is suggested by the fact that a change of psychophysical method or task often leads to distinctly different results. This is not to say that observers actually entertain many visual worlds interpretations explicitly, but merely that they don't necessarily resolve ambiguities when this is not specifically required for some action or decision.

Multiple Worlds (2012)



Consciousness: facing the consequences of decisions-in-action? (Schrodinger)

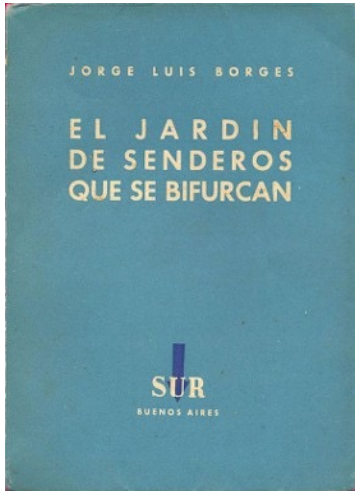
Do decisions have consequences when looking at art?

Non-representational art?

Kandinsky (1912)

I think that in a great many cases perceptions are more of the many visual worlds variety than like the single guess. You don't notice this in the laboratory because most psychophysical methods *force a unique response*. They simulate the decisions–in–action of daily life. That you don't notice the essential ambiguity of perception in real life is most likely due to the fact that you don't need to take decisions on issues on which no actions will be taken anyway. That the many visual worlds option is indeed likely is suggested by the fact that a change of psychophysical method or task often leads to distinctly different results. This is not to say that observers actually entertain many visual worlds interpretations explicitly, but merely that they don't necessarily resolve ambiguities when this is not specifically required for some action or decision.

Multiple Worlds (2012)



Consciousness: facing the consequences of decisions-in-action? (Schrodinger)

What is the nature of your visual awareness? A single visual world?

Joseph Mallord William Turner (Interior at Petworth c. 1837)