MIT Press • Open Encyclopedia of Cognitive Science

Unconscious Inference

Patrick Cavanagh[1]

[1]**Glendon College**

MIT Press

Published on: Sep 23, 2025

DOI: https://doi.org/10.21428/e2759450.3bef4aa1

Unconscious inference refers to the rapid and unconscious, high-level choices made by the visual system to quickly organize input from the eyes into a coherent scene comprising objects, depth, light, and action. The core challenge for vision is that each pattern in an image may have many interpretations. Dark areas may be shadows, holes, dark hair, or clothes. Making sense of the visual input requires inferences—good guesses based on regularities in visual scenes and knowledge about objects and scene properties. Critically, for the visual world to make sense fast enough, these inferences need to operate within about one-tenth to one-third of a second, prior to any conscious awareness of what will be seen. Vision is able to accomplish this remarkable feat because it occupies an enormous chunk of the brain, as much as 30% of the human cortex.

History

Hermann von Helmholtz (1924/2005) is often credited with the concept of unconscious inference, but Ibn al-Haytham (Sabra, 1989; for review Howard, 1996) preceded him by over 800 years. The inference processes, the choices made to explain the visual input, have sometimes been called midlevel or high-level vision (Blake & Sekuler, 2005; Gregory, 1980; Nakayama & Shimojo, 1992). They are part of a computational approach to cognition (here, visual cognition; Cavanagh, 2011) in which logical operations act on internal representations that stand for visual objects and scene elements. Although inferences do most of the complicated work of vision, there is very little yet known about how they operate. It is understood that inferences happen, and they are often described as rules or even laws. An appealing feature of *visual* inferences is that, due to the advances in the visual neurosciences compared to other fields, they may be easier to study. Once understood, these discoveries for visual operators can then be extended to similar logical operations that are duplicated across many levels of sensory and cognitive processing in the brain [see Hearing; Spatial Cognition]. This makes visual cognition a sort of in vivo laboratory preparation for all varieties of cognition.

Core concepts

Vision can be divided into two parts: measurement and inference (Marr, 1982). In the measurement stage, neurons with localized receptive fields indicate the presence of features ranging from brightness all the way to face identity (Quian Quiroga et al., 2023; Tsao et al., 2006). These processes and the neural architectures that realize them are quite well understood. The hard part comes next when inferences attempt to rapidly construct a representation of the scene that best accounts for these measurements.

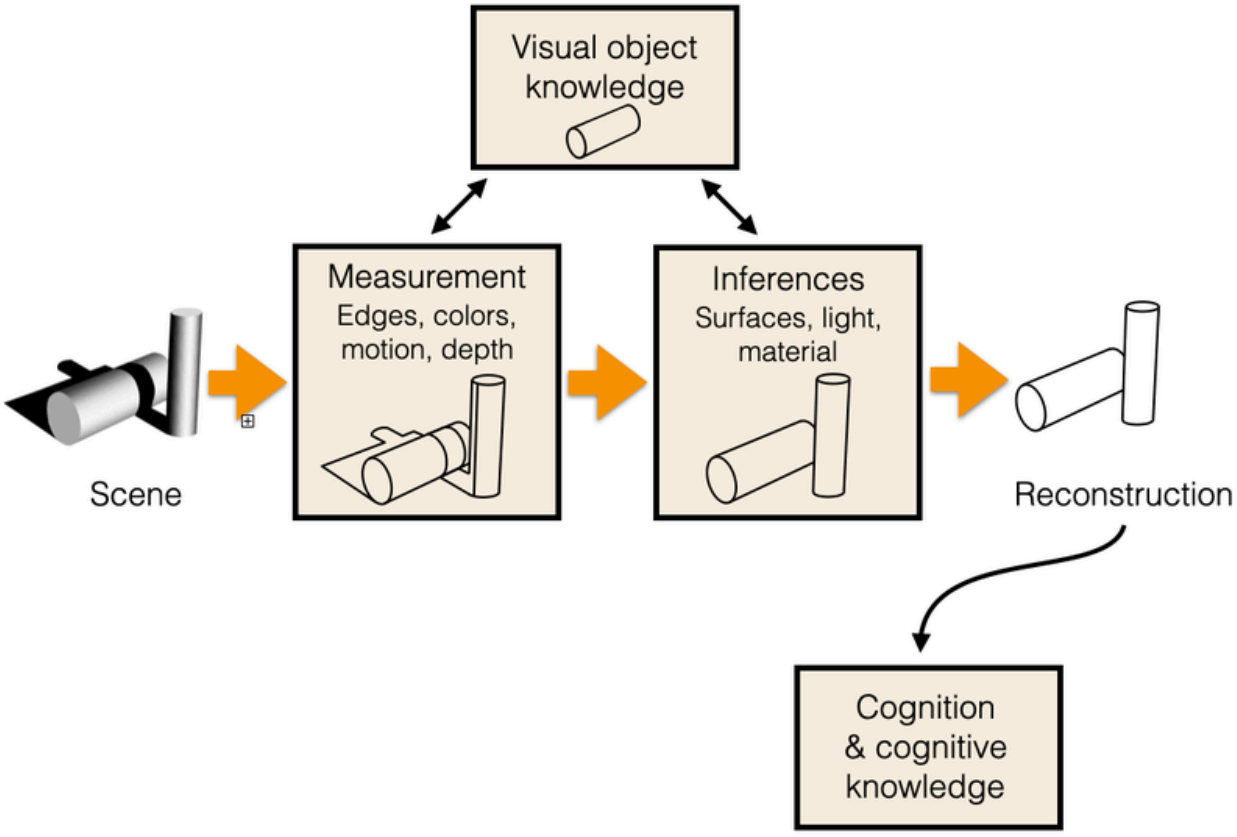

Figure 1
Visual processing begins with measurements of local features like orientation and color followed by an unconscious inference stage that rapidly makes the best story to explain these measurements. In both stages, "top-down" visual object knowledge may guide measurements and inferences. The output of the inference stage is a reconstruction of the scene populated by recognized objects, surfaces, light, and shadow. Conscious cognition may investigate the reconstruction based on an independent database of object knowledge but cannot influence the processing. When the expectations of cognition and perception disagree, it is seen as an illusion (**Cavanagh, 2024**). Adapted from **Casati and Cavanagh (2019)** with permission.

Inferences are generally based on specific, highly informative cues that are strong evidence for or against possible interpretations (Figure 1). For example, there may be a patch of texture that is typical of a grassy field, a dark region that may be shadow (Figure 2A), or a surface contour that disappears abruptly because one object passes behind another (Figure 2B). One approach to make sense of all cues is to use a voting scheme—a constraint satisfaction process (Cavanagh, 2011) that determines which interpretation is most consistent with all the inferences. Some of these informative cues and their inferences have been described as rules—for example, shadows must be darker (Casati & Cavanagh, 2019)—or laws such as the Gestalt law of continuity, in which two parts with aligned contours are taken to be one partially hidden object (Kellman & Shipley, 1991). Unlike actual laws, however, they can be outvoted by other competing inferences. It is critical that these visual inferences are fast enough to avoid reaching awareness. If all the busy work required for these inferences were conscious, the deliberations would not only slow down to

the speed of conscious processing but every perceptual outcome would also be drowned in the details of the cues and choices involved.

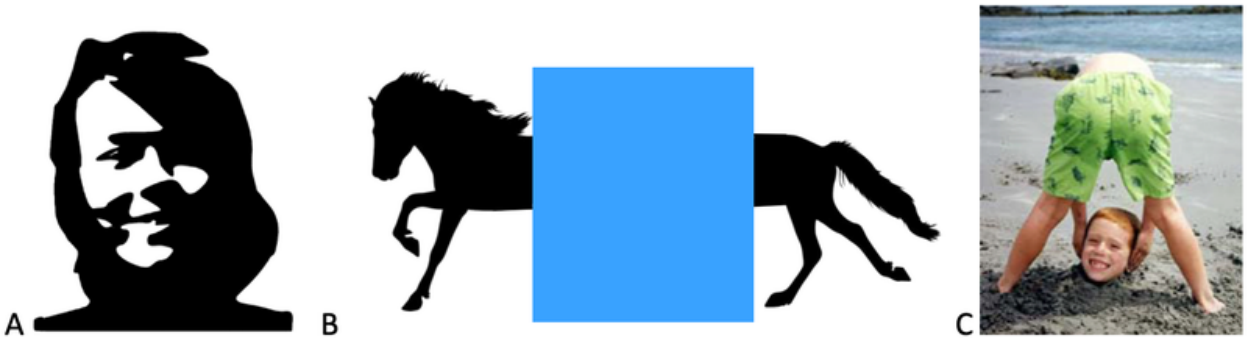

Figure 2

Object completion. (A) This high-contrast image has very little information, and yet it connects to visual object knowledge, and for many of us, this recovers the possible shape of a face. *(*B) Objects often occlude each other so that we see only partial views. However, they do not look partial. Our visual system infers that one object has hidden parts of the other. The abrupt ending of the horse's contours at the blue rectangle suggest that the rectangle is hiding parts of one or two horses. The continuity of the horse contours on either side of the rectangle supports the inference that the front and back end belong together in one very long horse. (C) The head on the sand seems to belong to the torso with the hidden head. The bias to complete the body schema initially overrules the unlikely separation of the body parts.

The goal of inference is to link the parts of the image together, so they can be recognized as distinct elements that have some spatial arrangement in the scene. This is also the main goal of vision, and this underlines the degree to which unconscious inferences are a central part of vision. For example, to combine an object's contours and surfaces (Figure 2B), its contours need to be linked to each other and to the objects to which they belong (Nakayama & Shimojo, 1992). This process may be realized in individual neurons that are selective for border ownership (Franken & Reynolds, 2021; Zhou et al., 2000), signaling that a contour belongs to the surface on one side of the contour (which then forms the object) and not the other (the background). The mutant horse of Figure 2B shows that the alignment of the back and belly contours across the intervening blue square leads to the impression that there is a single, stretched-out horse, despite experiences with horses that are never that long. The choices of which contours belong together are not always logical or compatible with previous experience.

Often inferences are based on a distinctive cue that is strongly linked to a particular object. An internal model of this object is then checked for supporting evidence in the image (e.g., Cavanagh, 1991; Lee & Mumford, 2003; Mumford, 1992); if the interpretation is correct, there should be other now-expected features present in the image. If it is a car, it should have wheels; if it is a person, it should have arms and legs. Once checked, the interpretation becomes the consciously experienced percept. In Figure 2C, the presence of legs and a torso leads an initial guess that this is a human body, and the verification stage checks for the rest of the body nearby. Interestingly, the person's head appears to have fallen in the sand below. These unexpected percepts, going far beyond what is specified, are evidence for the extensive

unconscious processing underlying perception. This loop of an initial guess or hypothesis followed by verification may fail, and then the next best interpretation must be considered, that is, there are two people here. Importantly, all this takes place unconsciously, within one-tenth to one-third of a second (Herzog et al., 2020; Hogendoorn, 2022). The unusual cases that remain unresolved during this preconscious phase—such as the misassigned head in Figure 2C—illustrate the existence of a checking stage by extending the process in time, allowing the subsequent revisions of the first interpretation to be available to conscious awareness.

In addition to piecing objects and surfaces together, vision must also deal with variations in lighting in different parts of the scene. If an area is dark, it may be a shadow or simply a dark surface in full light. If an area is light, it may be a light source itself or a reflection of a light source (a highlight) or a well-lit, white surface. Many inferences are required to decompose the light arriving at the eye into the surface properties of the object (color, lightness) and the intensity of light falling on it. Often the final choice is a compromise, as in Figure 3A, in which the dark area follows many of the "rules" for shadows (e.g., it is dark, and lies flat on the surface; Casati & Cavanagh, 2019), but it does break at least one—it has the wrong shape, as there is no key on the hook. In contrast, in Figure 3B, the shadow on the man's pants looks like a stain because it does not get linked to the object that is actually casting it (the hammer to its right). Without this ownership relation to its source, the inference that it is a shadow fails, leaving only a stain as an interpretation.

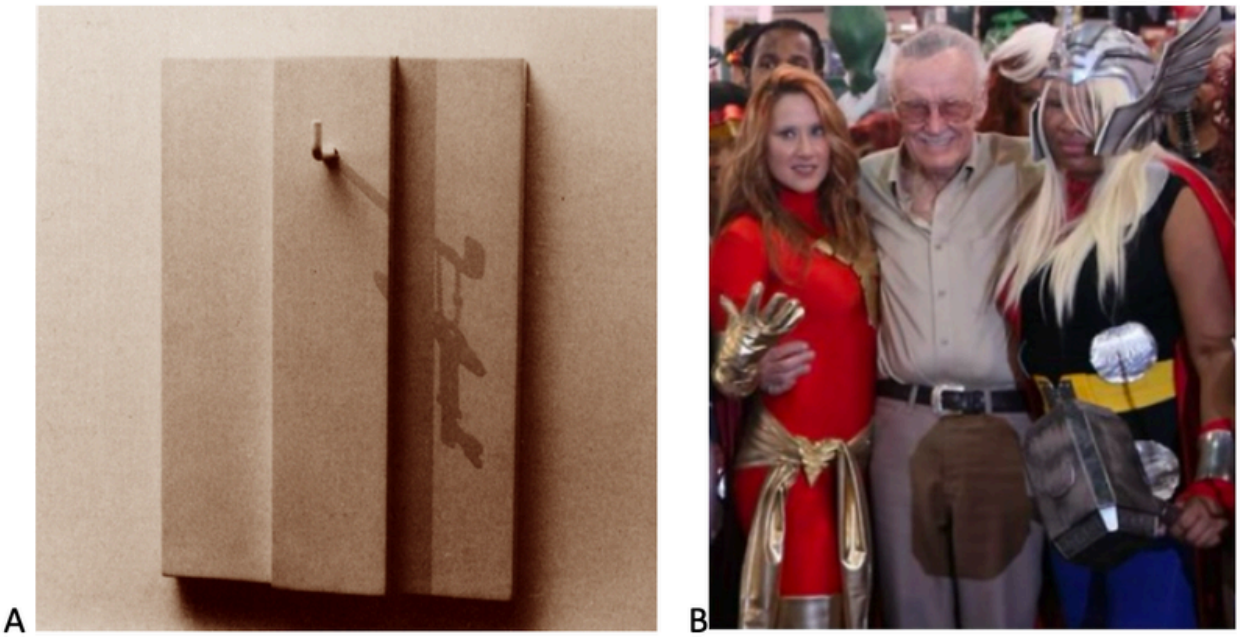

Figure 3

Shadows and reflections. (A) The shadow of the key is quite convincing, although there is no key. (B) When the shadow of Thor's hammer on the man's pants is not correctly linked to the hammer, it loses its shadow character and is interpreted as a stain.

Questions, controversies, and new developments

Inferences are often taken to be the domain of symbolic and conscious processing implemented in high-level cognitive brain regions (like frontal cortex) rather than in earlier visual areas. There is, however, compelling evidence for an independent visual intelligence that is separate from standard, everyday, reportable cognition. Specifically, our conscious knowledge does not affect our perception, as it is *cognitively impenetrable* (Pylyshyn, 1999); knowing that the two lines of the Müller-Lyer illusion have equal length, for instance, does not make them look any more equal. One remarkable example is the perception of causality that arises when one object hits another, knocking it away (Rolfs et al., 2013). Critically, this perception of causality is adaptable—exposure to long sequences of causal events makes subsequent tests appear less causal but only if the tests are at the same location on the retina as the adaptation; tests elsewhere show no effect. This retinal specificity is the signature of a purely visual process as opposed to a high-level, cognitive inference that would undoubtedly affect tests at all locations.

In contrast to computational, inferential approaches, others claim that perception is direct (Gibson, 1966, 1979; Raja, 2018). Specifically, the information in the scene is rich enough to specify appropriate percepts and actions that "resonate" to the scene layout and object identities, without any inferences or any form of internal representation. Nevertheless, there is as yet no compelling explanation of how this neobehavorist approach would be realized in the brain. Moreover, since direct perception is a reflexive connection from visual input to perception and action, it does not easily deal with alternatives such as those inherent in ambiguous and bistable images. The information available in the visual input is rarely sufficient to limit perception to a single interpretation; choices need to be made.

These choices must also be made by computer-based vision systems that have been developed for artificial intelligence, autonomous vehicle guidance, and face recognition [see Face Perception]. Many of these systems are based on architectures borrowed from the neural network properties of biological brains and are often referred to as deep neural nets. However, neural nets are not an alternative architecture—the visual system is a neural net. Unconscious inferences are just the stages and states in a neural net's processing that reach decisions about ambiguous input. The internal properties of the artificial neural net would most likely reveal computations equivalent to inferences. However, whether these would be unconscious inferences brings up a completely different question about machine awareness and the role of awareness in processing.

Broader connections

Each inference is an informed guess that selects the most likely interpretation based on expectations and visual input. This is also the description of Bayesian inferences, in which the probability of a particular guess or hypothesis can be determined based on prior beliefs and observed data [see Bayesian Models of Cognition; Bayesianism]. However, the Bayesian approach only evaluates hypotheses; it does not generate them. It selects the most likely among an existing set of hypotheses, a ranked table lookup approach. In contrast, the core function of unconscious visual inferences is the generation of guesses and hypotheses,

including completely unexpected ones, that can be suboptimal in the Bayesian sense (Anderson et al., 2011).

The unconscious inferences require verification against additional image features, and this recurrent feedforward, feedback loop brings the inference process into the domain of predictive processing [see The Free Energy Principle]. In this framework, originally proposed as an analysis by synthesis process (Lee & Mumford, 2003; Yuille & Kersten, 2006), the internal model of the scene generates a prediction of what the input should be. Any discrepancy between the prediction and the input is used to update the model. This is again an evaluation process that does not specify the mechanisms that generate the original internal model. On the other hand, one form of deep neural nets, autoencoders (Hinton, 2007), does act to match activity reconstructed from higher-level representations to the actual input. They are an instantiation of the essential loop of predictive coding and have been used to model human color and material perception (Fleming & Storrs, 2019; Li et al., 2022) and object-based attention (Al-Tahan & Mohsenzadeh, 2021; Cavanagh, 2023), among other visual processes. The neural net architecture of autoencoders does allow the learning of object and scene properties that can support the creation of an internal model of a scene, and they offer an attractive, alternative approach for studying the visual inference processes that lead to these internal models.

Unconscious inferences represent a sophisticated level of processing that constructs the perceptual experience of the visual world. They are fast enough to remain unconscious and so avoid cluttering the perceptual experience with the details of the decision processes. Of course, inferences are just best guesses and will occasionally lead to incorrect representations. When this happens, it may be seen as an illusion, a magician's trick, or the occasional misperception. Nevertheless, the error rate is remarkably low, and the scene reaching awareness within a tenth to a third of a second is, reassuringly, quite accurate.

Acknowledgments

This research was supported in part by grants from the Natural Sciences and Engineering Research Council of Canada grant RGPIN-2019-03938. Many thanks to Josée Rivest and Stuart Anstis for comments on this article.

Further reading

- Cavanagh, P. (2011). Visual cognition. *Vision Research*, *51*(13), 1538-1551. https://doi.org/10.1016/j.visres.2011.01.015
- Gregory, R. L. (1980). Perceptions as hypotheses. *Philosophical Transactions of the Royal Society B*, *290*(1038), 181–197. https://doi.org/10.1098/rstb.1980.0090
- Gregory, R. L. (2015). *Eye and brain: The psychology of seeing* (5th ed.). Princeton University Press

References

- Al-Tahan, H., & Mohsenzadeh, Y. (2021). Reconstructing feedback representations in the ventral visual pathway with a generative adversarial autoencoder. *PLoS Computational Biology*, *17*(3), e1008775. https://doi.org/10.1371/journal.pcbi.1008775 ↩
- Anderson, B. L., O'Vari, J., & Barth, H. (2011). Non-Bayesian contour synthesis. *Current Biology*, *21*(6), 492-496. https://doi.org/10.1016/j.cub.2011.02.011 ↩
- Blake, R., & Sekuler, R. (2005). *Perception* (5th ed.). McGraw Hill. ↩
- Casati, R., & Cavanagh, P. (2019). *The visual world of shadows*. MIT Press. ↩
- Cavanagh, P. (1991). What's up in top-down processing? In A. Gorea (Ed.), *Representations of vision: Trends and tacit assumptions in vision research* (pp. 295-304). Cambridge University Press. ↩
- Cavanagh, P. (2011). Visual cognition. *Vision Research*, *51*(13), 1538-1551. https://doi.org/10.1016/j.visres.2011.01.015 ↩
- Cavanagh, P., Caplovitz, G. P., Lytchecnko, T. K., Maechler. M. R., Tse, P. U., & Sheinberg, D. R. (2023). The architecture of object-based attention. *Psychonomic Bulletin & Review*, *30*, 1643–1667. https://doi.org/10.3758/s13423-023-02281-7 ↩
- Fleming, R. W., & Storrs, K. R. (2019). Learning to see stuff. *Current Opinion in Behavioral Sciences*, *30*, 100–108. https://doi.org/10.1016/j.cobeha.2019.07.004 ↩
- Franken, T. P., & Reynolds, J. H. (2021). Columnar processing of border ownership in primate visual cortex. *Elife*, *10*, e72573. https://doi.org/10.7554/eLife.72573 ↩
- Gibson, J. J. (1966). *The senses considered as perceptual systems*. Houghton Mifflin. ↩
- Gibson, J. J. (1979). *The ecological approach to visual perception.* Houghton Mifflin. ↩

- Gregory, R. L. (1980). Perceptions as hypotheses. *Philosophical Transactions of the Royal Society B*, *290*(1038), 181–197. https://doi.org/10.1098/rstb.1980.0090 ↩
- Helmholtz, H. V. (2005). *Treatise on physiological optics (vol. 3, J. P. C. Southall, Trans.)*. Dover. (Original work published 1924) ↩
- Herzog, M. H., Drissi-Daoudi, L., & Doerig, A. (2020). All in good time: Long-lasting postdictive effects reveal discrete perception. *Trends in Cognitive Sciences*, *24*(10), 826-837. https://doi.org/10.1016/j.tics.2020.07.001 ↩
- Hinton, G. E. (2007). Learning multiple layers of representation. *Trends in Cognitive Sciences*, *11*(10), 428–434. https://doi.org/10.1016/j.tics.2007.09.004 ↩
- Hogendoorn, H. (2022). Perception in real-time: Predicting the present, reconstructing the past. *Trends in Cognitive Sciences*, *26*(2), 128-141. https://doi.org/10.1016/j.tics.2021.11.003 ↩
- Howard, I. P. (1996). Alhazen's neglected discoveries of visual phenomena. *Perception*, *25*(10), 1203–1217. https://doi.org/10.1068/p251203 ↩
- Kellman, P. J., & Shipley, T. F. (1991). A theory of visual interpolation in object perception. *Cognitive Psychology*, *23*(2), 141-221. https://doi.org/10.1016/0010-0285(91)90009-d ↩
- Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America A*, *20*(7), 1434-1448. https://doi.org/10.1364/josaa.20.001434 ↩
- Li, Q., Gomez-Villa, A., Bertalmío, M., & Malo, J. (2022). Contrast sensitivity functions in autoencoders. *Journal of Vision*, *22*(6), 8. https://doi.org/10.1167/jov.22.6.8 ↩
- Marr, D. (1982). *Vision*. W. H. Freeman. ↩

- Mumford, D. (1992). On the computational architecture of the neocortex. II. The role of cortico-cortical loops. *Biological Cybernetics*, *66*(3), 241-251. https://doi.org/10.1007/BF00198477 ↩
- Nakayama, K., & Shimojo, S. (1992). Experiencing and perceiving visual surfaces. *Science*, *257*(5075), 1357–1363. https://doi.org/10.1126/science.1529336. ↩
- Pylyshyn, Z. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences*, *22*(3), 341-365. https://doi.org/10.1017/s0140525x99002022 ↩
- Quian Quiroga, R., Boscaglia, M., Jonas, J., Rey, H. G., Yan, X., Maillard, L., Colnat-Coulbois, S., & Rossion, B. (2023). Single neuron responses underlying face recognition in the human midfusiform face-selective cortex. *Nature Communications*, *14*(1), 5661. https://doi.org/10.1038/s41467-023-41323-5 ↩
- Raja, V. (2018). A theory of resonance: Towards an ecological cognitive architecture. *Minds and Machines*, 28(1), 29-51. https://doi.org/10.1007/s11023-017-9431-8 ↩
- Rolfs, M., Dambacher, M., & Cavanagh, P. (2013). Visual adaptation of the perception of causality. *Current Biology*, *23*(3), 250-254. https://doi.org/10.1016/j.cub.2012.12.017 ↩
- Sabra, A. I. (Ed.). (1989). *The optics of Ibn Al-Haytham: Books I–III*: On direct vision (A. I. Sabra, Trans.). The Warburg Institute. ↩
- Tsao, D. Y., Freiwald, W. A., Tootell, R. B., & Livingstone, M. S. (2006). A cortical region consisting entirely of face-selective cells. *Science*, *311*(5761), 670-674. https://doi.org/10.1126/science.1119983 ↩
- Yuille, A., & Kersten, D. (2006). Vision as Bayesian inference: Analysis by synthesis? *Trends in Cognitive Sciences*, *10*(7), 301-308. https://doi.org/10.1016/j.tics.2006.05.002 ↩
- Zhou, H., Friedman, H. S., & Von Der Heydt, R. (2000). Coding of border ownership in monkey visual cortex. *Journal of Neuroscience*, *20*(17), 6594-6611. https://doi.org/10.1523/JNEUROSCI.20-17-06594.2000 ↩