Understanding vision in wholly empirical terms

Dale Purves[a,1], William T. Wojtach[b], and R. Beau Lotto[c]

[a]Center for Cognitive Neuroscience, Department of Neurobiology, Duke University, Neuroscience and Behavioral Disorders Program, Duke–National University of Singapore Graduate Medical School, Singapore 169857; [b]Center for Cognitive Neuroscience, Duke University, Levine Science Research Center, Durham, NC 27708; and [c]Institute of Ophthalmology, University College London, London EC1V 9EL, United Kingdom

This article considers visual perception, the nature of the information on which perceptions seem to be based, and the implications of a wholly empirical concept of perception and sensory processing for vision science. Evidence from studies of lightness, brightness, color, form, and motion all indicate that, because the visual system cannot access the physical world by means of retinal light patterns as such, what we see cannot and does not represent the actual properties of objects or images. The phenomenology of visual perceptions can be explained, however, in terms of empirical associations that link images whose meanings are inherently undetermined to their behavioral significance. Vision in these terms requires fundamentally different concepts of what we see, why, and how the visual system operates.

Bayes' theorem | empirical ranking | illusion | perceptual qualities | visual context

A major obstacle in explaining how the visual system generates useful perceptions is that light stimuli cannot specify the objects and conditions in the world that caused them. This quandary, called the inverse optics problem, pertains to every aspect of visual perception (Fig. 1). The evidence described in this article all points to the same basic strategy for contending with this biological challenge: visual perceptions arise by linking retinal stimuli with useful behaviors according to feedback from trial and error interactions with a physical world that cannot be revealed by sensory information. In consequence, what we see is a world determined by the behavioral significance of retinal images for the species and the perceiver in the past rather than by an analysis of stimulus features in the present.

Since the late 1950s, the focus of visual neuroscience has been on the receptive field properties of neurons in the primary and higher-order visual pathways in experimental animals (1, 2). Implicit in this research is the idea that perception arises from neural mechanisms that detect features and filter the characteristics of retinal stimuli, analyze those features hierarchically and in parallel, and eventually recombine this information in visual cortex to represent external reality. Given the quandary posed by the inverse optics problem, however, this paradigm is not feasible. Here, we describe evidence that the visual system relies on a fundamentally different strategy.

A Wholly Empirical Concept of Vision

To understand how vision can contend with the inverse problem, consider how a player could behave successfully in a game of dice when direct analysis (e.g., weighing the dice, taking them apart, etc.) is precluded as a way to determine whether the dice are fair or loaded. This scenario presents a problem similar to the one confronting vision, where direct information about the physical world is likewise unavailable.

Despite the exclusion of direct analysis, an operational evaluation of the dice and the implications for successful behavior (betting advantageously) can nonetheless be made by tallying the frequency of occurrence of the numbers that come up over the course of many throws. Just as the properties of the physical world determine the frequency of occurrence of different retinal stimuli, the physical properties of the dice determine the frequency of occurrence of the different numbers rolled. If the numbers come up about equally, then a player should behave as if the dice are fair and bet accordingly; conversely, if some numbers appear more frequently than others, then the dice are loaded and to succeed, the player's betting strategy should change. Note that, in this way of contending with the problem, the physical properties of the dice are not represented.

Because two dice are involved, operational evaluation can improve performance still further by taking advantage of the fact that the numbers on one die are relevant to the numbers on the other. If a particular number on one of the dice is taken as a "target" then tallying how often that number comes up together with each of the possible "context" numbers on the other die provides a more refined guide to betting behavior. For example, if a 5 on the target die were often associated with a 2 on the other contextual die, then betting on a total of 7 would be a good strategy. The more frequently a target number on one die is associated with a context number on the other, the greater the chance that a given behavioral response predicated on that combination will succeed.

The human visual environment is like loaded dice in that the complex physical properties of the world ensure that stimuli do not come up equally over time. By progressively modifying the connectivity of the brain according to the frequency of occurrence of the targets and contexts in stimuli that the visual environment provides, appropriate visual behavior can be generated, despite the inverse problem. Over time, useful visually guided behaviors can be generated even though the physical properties that gave rise to retinal stimuli are not available in images as such. In consequence, any attempt to explain vision in terms of image analysis is not viable (Fig. 1).

Rationalizing vision based on the target–context associations instantiated in the visual system according to behavioral responses to retinal stimuli rather than the properties of objects themselves, however, predicts that visual perceptions will not correspond to the actual properties of the physical world (just as the associations among the numbers on the faces of each die do not correspond to the physical properties of the dice). In this framework, then, the basis for what we see is not the physical qualities of objects or actual conditions in the world but operationally determined perceptions that promote behaviors that worked in the past and are thus likely to work in response to current retinal stimuli. Although no one would argue with the idea that vision evolved to promote useful behavior, the theory presented here relies entirely on the history of behavioral success rather than on image analysis. We, therefore, refer to this strategy of vision as "wholly empirical."
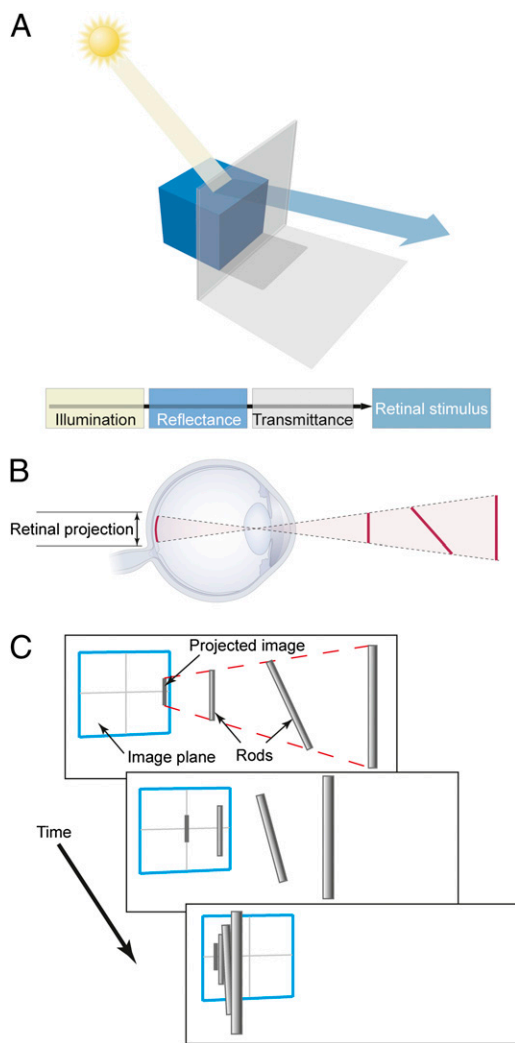
Fig. 1. The inverse optics problem. (*A*) The conflation of illumination, reflectance, and transmittance in retinal images. Many combinations of these physical characteristics of the world can generate the same retinal stimulus. (*B*) The conflation of physical geometry in images. The same image can be generated by objects of different sizes, at different distances from the observer, and in different orientations. (*C*) The conflation of speed and direction in images of moving objects. The same projected motion on the retina can be generated by different objects with various orientations moving in different directions and at different speeds in the physical world.

Testing the Theory

The merits of a wholly empirical strategy of vision can be assessed by using the frequency of occurrence of stimuli to stand in for the trial and error experience in the human visual environment that would have linked retinal images to useful visually guided behaviors, thus contending with the inverse problem. The argument for this approach is much the same as the loaded dice analogy: by tallying up the frequency of occurrence of different targets and contexts in visual stimuli generated in nature and instantiating this information in visual system connectivity, perceptions arising on this basis should be predictable. The evidence supporting a wholly empirical theory depends on whether such predictions accord with perceptual experience.

The following sections provide examples of how this approach works for three different perceptual qualities: lightness and brightness, form, and motion (ref. 3 provides a full account of how these and other basic visual qualities can be understood in empirical terms).

Lightness and Brightness. The term "lightness" refers to the lighter or darker appearance of objects because of the amount of light that a surface returns to the eye; "brightness" refers to the brighter or dimmer appearance of objects that are themselves sources of light (e.g., the sun, fire, and light bulbs). Like all other perceptual qualities, lightness and brightness are not subject to direct measurement and can be evaluated only by asking observers to report what they see.

Because increasing the luminance of a stimulus increases the number of photons captured by photoreceptors and thus the relative action potential output of the retina at any given level of ambient light, a simple expectation would be that physical measurements of light intensity and perceptions of lightness and brightness are directly proportional. A corollary is that objects in a scene returning the same amount of light to the eye should appear equally light or bright. Perceptions of lightness and brightness, however, fail to meet these expectations. For example, two patches having the same luminance are perceived as being differently light or bright when their backgrounds have different luminance values (Fig. 2*A*). Moreover, as has long been known (4, 5), presenting luminance targets in more complex contexts makes these effects much stronger (Fig. 2*B*).

Rather than trying to explain these perceptual phenomena in terms of analyzing and processing information in the retinal stimulus, an empirical framework depends solely on visual history: natural selection would have instantiated random changes in the structure and function of the visual systems of ancestral forms according to how well the ensuing behavior served reproductive success. Any configuration of neural circuitry that better served the empirical link between visual stimuli and successful behavior would tend to increase among the members of the species, whereas less useful circuit configurations would not. Similarly, experience-dependent refinements of connectivity during postnatal development and adult life would allow individuals to contend with the challenge presented by the inverse problem more successfully than could be achieved on the basis of inherited circuitry alone.

To test the validity of this idea, templates similar to the patterns in Fig. 2*A* were used to sample a database of images arising from natural scenes; this method allows a determination of the frequency of occurrence of different target luminance values in particular luminance surrounds. These data, therefore, indicate how often a target luminance embedded in a surrounding context occurs in natural stimuli—the information that historically successful behavior would gradually have encoded in the visual system to contend with the inverse problem. Thus, the identical targets in Fig. 2*A* look differently light or bright because the frequency of occurrence of the retinal projections generated by natural sources is different.

To take a specific example, the targets (T) with different luminance values in Fig. 3*A* indicate possible target luminance values in natural scenes in which the targets are surrounded by differently luminant contexts. The probability distribution of target luminance values co-occurring with the luminance patterns of two different contexts is shown in Fig. 3*B*. As might be expected, the values of the targets occurring most often are similar to the mean luminance of the corresponding surrounds. To predict the perceived lightness or brightness elicited by target luminance values on wholly empirical grounds, the distribution of co-occurring target and surrounding luminance values can be expressed in cumulative form to define the summed probability that a particular target luminance will have occurred in a given surround (Fig. 3*C*). Thus, the perception of lightness or brightness is predicted by the relative percentile rank (probability × 100%) of each co-occurring target and surround in these cumulative distributions, with a higher rank corresponding to a greater perceived lightness or brightness relative to the surround. For example, in Fig. 3*C*, the cumulative probability dis-

Fig. 2. Discrepancies between luminance and perceptions of lightness and brightness. (*A*) Standard demonstration of the simultaneous brightness contrast effect. A target (the diamond) on a less luminant background (*Left*) is perceived as being lighter or brighter than the same target on a more luminant background (*Right*), even though the two targets have the same measured luminance; if both targets are presented on the same background, they appear to have the same lightness or brightness (*Inset*). (*B*) Although the gray target patches again have the same luminance and appear the same in a neutral setting (*Inset*), a striking perceptual difference in lightness is generated by contextual information that one set of patches is in shadow (those on the riser of the step) and another set in light (those on the surface of the step). [Reprinted with permission from ref. 6 (Copyright 2003, Sinauer Associates)].

tributions are given by $P(<L_T> | L_S)$, where $<L_T>$ represents the full range of the possible luminance values of a target patch that have occurred in conjunction with the specific luminance value of the surround (L_S). The distribution $P(<L_T> | L_S)$, therefore, provides an empirical scale that ranks all the target luminance values that have co-occurred with the surround in question: among all of the physical sources of a target patch that have co-occurred with the surround in human experience, the distribution indicates the percentage that generated target patches having luminance values less than T and the percentage that generated target patches with luminance values greater than T.

Fig. 3. Predicting the perception of lightness or brightness by the frequency of target and context luminance values in stimuli generated by the world [explanation in the text; adapted from Yang and Purves (7)].

For any particular target luminance (T) along the abscissa, the corresponding rank along the ordinate (T*) predicts the relative perception of the target. As shown in Fig. 3*C*, the higher that T* ranks on this empirical scale, the lighter or brighter the perception of that particular patch.

Without a context, predicting the lightness or brightness elicited by different luminance values according to their percentile rank would only have shown that a higher luminance value would always have a higher rank than a lower value. However, once context is introduced, which it always is in natural stimuli, the relationship between a luminance value and its rank is no longer a simple one. Because the empirical scales associated with various surrounds can be quite different, the same target luminance value can have very different rankings when embedded in different contexts and thus can generate quite different perceptions of lightness or brightness, as is apparent in Fig. 2.

It follows from this strategy of vision that calling any perception of lightness or brightness a "visual illusion" is incorrect. Rather, the perceptions that arise are simply the signature of how the visual brain generates *all* subjective responses to luminance. In these terms, then, the conventional distinction between veridical and illusory perception is false; by the same token, making inferences about the physical properties of objects or states of the world is not how vision seems to work. Notice as well that, in this framework, experience with a series of images as such is insufficient to generate useful visual perceptions. Because

the mechanisms for instantiating the needed visual circuitry are natural selection and activity-dependent neural plasticity, feedback from the success or failure of ensuing behaviors is essential to complete the biological loop underlying a wholly empirical visual strategy.

Finally, we emphasize that any counterargument maintaining that the perceptions elicited by the stimulus in Fig. 3A could also be explained by a visual strategy that statistically assesses the physical properties of objects (their surface reflectance values in this case) and the relative probabilities of their co-occurrence is undermined by the fact that the visual system cannot make such determinations. Even if predictions from such models were to conform to some set of psychophysical data, they would fail to meet a necessary criterion for any explanation of visual perception (i.e., contending with the fact that the inverse problem prevents direct access to the environment).

Form. In considering the merits of a wholly empirical theory of vision, plausibly accounting for a single phenomenon or even a category of phenomena like perceptions of lightness and brightness will not suffice. To be considered an explanation of vision, this or any theory must account for a wide range of perceptions within the context of the inverse problem.

Perceived geometrical form is another broad category that can be examined in empirical terms (3, 8). In addition to conflating the physical parameters that determine the quantity and quality of light reaching the eye from objects and conditions in the world, the parameters that define the location and arrangement of the objects in space—their size, distance, and orientation—are also inextricably intertwined in retinal images (Fig. 1B). Thus, the inverse problem is as much an obstacle in generating useful perceptions of spatial relationships conveyed by light as it is in generating perceptions of lightness or brightness from luminance values returned to the eye, since the spatial characteristics of physical sources cannot be determined by any operation on images as such.

Many discrepancies between measured and perceived geometry have been described, and attempts to rationalize these effects have a long history (3, 8, 9). If vision operates empirically, then these discrepancies should also be explainable in terms of the frequency of occurrence of 2D images generated by 3D sources and behavioral feedback from operating in the human visual environment. In this case, the database that serves as a proxy for human experience is acquired by laser range scanning of natural scenes. This method relates the geometries in projected images with their underlying sources in the world, allowing the frequency of occurrence of geometrical forms on the retina to be measured. These data can then be used to predict the geometries that people should see if vision is empirically determined.

A simple example of how such information can explain the perception of form is how we see the length of lines. In the absence of other contextual information, a line drawn on a piece of paper or computer screen would be expected to correspond more or less directly to the length of the line projected onto the retina. If perception were determined by the detection and processing of image features, then the perceived length of a line should be directly related to its projected length. As it turns out, however, vertical lines look 10–15% longer than horizontal lines (10–14). Even stranger, the perception of line length varies continuously as a function of orientation, with the maximum apparent length being elicited by a line stimulus oriented about 30° from vertical (Fig. 4).

To explain these perceptions in empirical terms, a first step is to determine from a database how often in the past a line with a given length and orientation on the retina has been experienced by human observers. In an empirical framework, the apparent lengths of real-world lines elicited by projected lines in different orientations should be predicted by the frequency of occurrence of projected line lengths (length being the target in
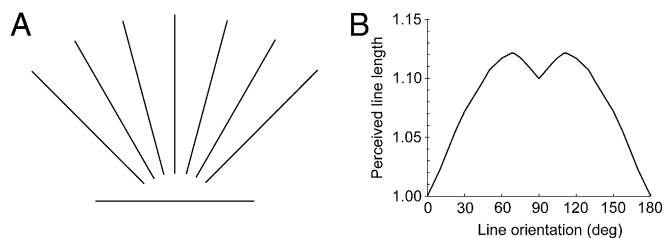

Fig. 4. Variation in apparent line length as a function of orientation. (A) The horizontal line in this figure looks somewhat shorter than the vertical or oblique lines, despite the fact that all of the lines have the same measured length. (B) The apparent length of a line reported by subjects as a function of its orientation in the retinal image (expressed as the angle between the line and the horizontal axis). The maximum length seen by observers occurs when the line is oriented ~30° from vertical, at which point it appears about 10–15% longer than the minimum length seen when the orientation of the stimulus is horizontal. The data shown here is an average of psychophysical results reported in the literature (15, 16).

this case) as a function of orientation (the context). This information can be extracted from the database by tallying the frequency of occurrence of projected straight lines in different orientations arising from geometrically straight lines in the world (Fig. 5A). Each of the probability distributions derived in this way can then be expressed in cumulative form to provide an empirical scale of the frequency of occurrence of line lengths projected at specific orientations (Fig. 5B). As in predicting lightness or brightness perceptions, this information can be expressed in terms of the percentile rank of the projected length of a given line in relation to *all* other projected lines in that orientation. The rank of any line in the cumulative probability distributions indicates what percentage of projected lines in a given orientation have been shorter than the line in question over the course of human experience, and what percentage of projected lines have been longer. As illustrated in Fig. 5B, the percentile rank for projected lines of any given length varies as a function of orientation.

If a wholly empirical strategy is correct, the apparent length of lines as a function of their projected orientation on the retina (Fig. 4B) should be predicted by the percentile rank of a given projected line as its orientation changes. Analysis of the natural scene database shows that, for lines near vertical, a greater number of shorter than longer projected lines have occurred in the sum of experience compared with horizontal line projections of the same length (Fig. 5B). This fact means that the percentile rank of a vertical line of any length on the retina will be higher, and thus be seen as longer, relative to a horizontal line of the same length. As indicated in Fig. 5C, the shape of the function defined by the percentile rankings of a line on the retina in different orientations closely follows the psychophysical function of perceived length as a function of orientation (Fig. 4B).

Notice that these perceived line lengths do not arise because there are more physically longer vertical objects in the environment than horizontal ones; in fact, longer horizontal objects are more frequent (chapter 3 in ref. 8). Rather, the different perceptions of length elicited by the same physical line in different orientations are again an indication of how the visual system contends with the inverse problem. Consequently, the relative frequencies of occurrence of projected lines in different orientations should predict perceptions of length, which they do.

Motion. The perception of motion provides another category that must be explained in empirical terms. Physical motion refers to the speed and direction of objects as they change location in the world. In perception, however, motion is defined subjectively and must again contend with the inverse problem: neither the speed
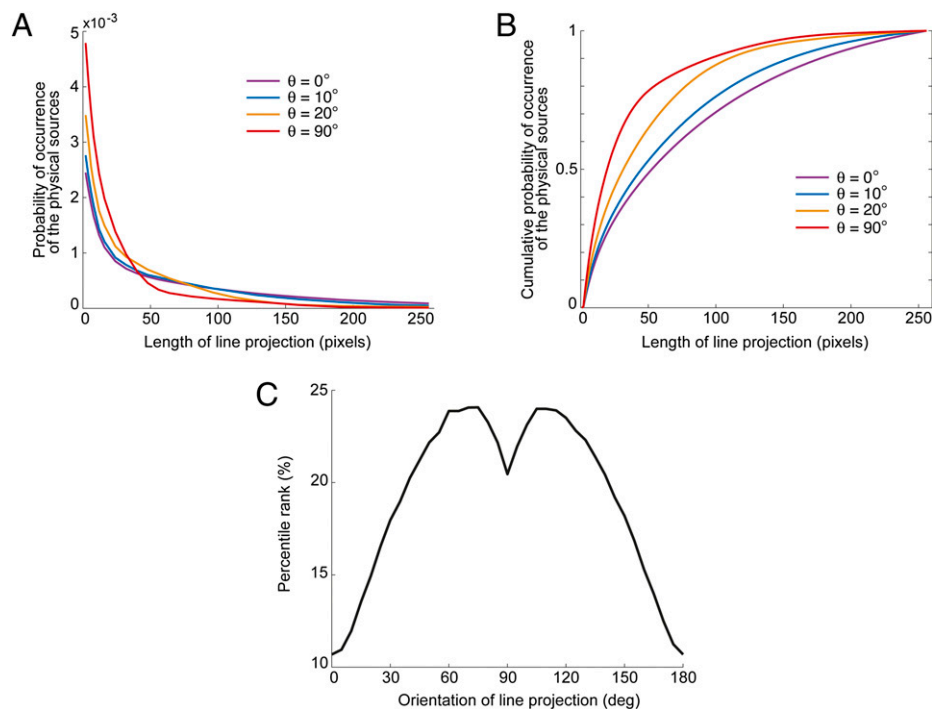
Fig. 5. The frequency of occurrence of lines projected onto the retina in different orientations determined by laser range scanning of typical environments. (*A*) Probability of the physical sources capable of generating lines of different lengths and orientations (θ) in the retinal image. (*B*) Cumulative probability distributions calculated from the distributions in *A*. The cumulative values for any given point on the abscissa are obtained by calculating the area underneath the curves in *A* that lie to the left of a line of that length in the relevant distribution. This value indicates how many lines in that orientation in retinal images have been shorter than the projected line in question and how many have been longer. (*C*) The predicted function for projected lines 6 pixels in length in different orientations derived from the cumulative probability distribution in *B*. The predicted function in *C* is similar to the psychophysical function of perceived line lengths in Fig. 4*B* (15, 16).

nor the direction of moving objects is specified by motion in 2D retinal projections (Fig. 1*C*). Vision scientists have long sought to explain motion perception in terms of the receptive field properties of visual neurons (17–20). At the same time, psychologists have tried to account for apparent speed and direction in terms of various ad hoc models (18, 21–25). A different conception of the problem consistent with the explanations of lightness, brightness, and geometry is that perceived motion is also determined historically by the frequency of occurrence of motion in retinal images and feedback from behavior in the human visual environment.

There are many puzzling motion percepts. Preeminent among these percepts are the odd way that the speed of motion is seen (e.g., the flash-lag effect) (Fig. 6*A*) (22, 23, 25, 26) and the dramatic changes in the apparent direction of motion that occur when a moving object projects to the eye through an aperture (27). These discrepancies between physical and perceived motion present another problem for the evolution and development of useful vision: observers must respond to the speeds and directions of objects in 3D space, but this information is not available in the 2D speeds and directions projected on the retina.

For example, if perceived speed is determined empirically, then the lag reported by observers for different object speeds shown in Fig. 6*B* should be predicted in much the same way as perceived line length in the previous section. Because any image speed can be generated by a wide range of object speeds (Fig. 1*C*), perceived speed should correspond to the frequency of occurrence of projected speeds arising from objects in the environment rather than to the actual speeds of objects. By converting this probability distribution of projected image speeds into cumulative form, the summed probability that moving objects undergoing perspective transformation have given rise to

particular image speeds can be determined. If the flash-lag effect is a signature of an empirical strategy of visual motion processing, then the lag reported by observers for the different image speeds in Fig. 6*B* should be predicted by the relative positions of image speeds in the cumulative probability distribution. As in predicting lightness, brightness, and form, a database—in this instance, projections of moving objects in a virtual environment—serves as a proxy for the frequency of occurrence of different image speeds generated by moving objects in past human experience (28).

The predictions made from the motion database correspond to the percentile rank (probability × 100%) of each image speed, with higher rankings indicating perceptions of greater speed. Because a stationary flash has an image speed of 0°/s (and therefore, a percentile rank of 0%), any moving image (the target) relative to a flash (the context) will have a higher ranking; consequently, a flash should always be perceived to lag behind a moving object. Moreover, because an increase in image speed corresponds to a higher ranking, the perceived magnitude of the flash-lag effect should also increase according to the nonlinear function of Fig. 6*B*. As shown in Fig. 6*C*, a direct correlation of the psychophysical results with the percentile ranking of image speeds predicts perceived lag quite well (see also ref. 29). The differences in the probability distributions generated by the geometry of the world also predict the variable directions of motion people see when the light reflected from objects is projected through apertures (30).

Discussion

Describing What We See on an Empirical Basis. If the physical properties of objects in the world are not represented in vision, then what is the proper way to describe what we see? An in-
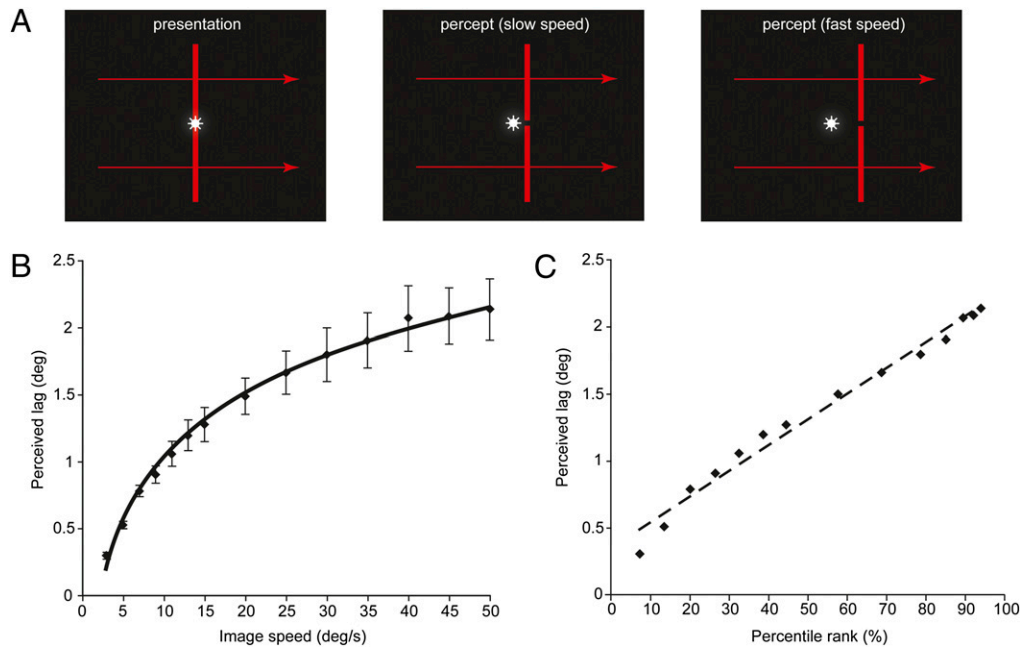
Fig. 6. Discrepancies between physical and perceived speeds. (*A*) The flash-lag effect. When a flash of light (asterisk) is presented in alignment with a moving object (the red bar; *Left*), the flash is seen lagging behind the position of the object (*Center*). The apparent lag increases as the speed of the moving object increases (*Right*). The amount of lag as a function of object speed can be determined by asking subjects to align the flash with their perception of the moving stimulus at various object speeds. (*B*) Psychophysical function describing the flash-lag effect for image speeds up to 50°/s. The curve is a logarithmic fit to the lag reported by 10 observers. Bars are ±1 SEM. (*C*) Plotting the perceived lag reported by observers in *B* against the percentile ranking of image speeds from the motion database illustrates the correlation between both sets of data. The deviation from a linear fit (dashed line) indicates that >97% of the observed data are accounted for on an empirical basis [adapted from Wojtach et al. (28)].

structive example is the sensory quality of pain. It is clear that pain does not correspond to the physical objects or conditions that cause it; rather, pain is a subjective quality that we evolved to sense because such perceptions promote survival. In an empirical framework, visual qualities bear this same relationship to the world. Although we routinely attribute visual perceptual qualities to objects and environmental conditions, our experiences of lightness, brightness, color, form, and motion are likewise subjective qualities that simply promote useful behavior. Accordingly, it would be best to describe visual perceptions in terms similar to those used to describe pain, for which the concept of representation makes no sense. Visual perceptions, like the perception of pain, do not stand for the properties of objects in the physical world, although the world, of course, generates the relevant stimuli.

The distinction between visual perception conceived in terms of the awareness of behaviorally useful qualities vs. conceiving perception in terms of representations of the physical world would be a philosophical point only, were it not for the associated neurobiological implications. If vision does not represent the properties of objects and conditions in the world, then neither do its underlying anatomical and physiological mechanisms, which must, therefore, be thought of, examined, and tested in different terms.

Mechanisms. The biological mechanisms underlying a wholly empirical theory of vision are worth restating. The driving force that instantiates the links between light stimuli and visual behaviors during the evolution of a visual species is natural selection: random changes in the structure and function of the visual systems of ancestral forms have persisted—or not—in descendants according to how well they serve the reproductive success of the animal whose brain harbors that variant. Any configuration of neural circuitry that mediates more successful responses to visual stimuli will increase among the members of

the species, whereas less useful circuit configurations will not. The significance of this conventional statement about the phylogeny of any biological system is simply the existence of a universally accepted mechanism for instantiating and updating neural associations between light stimuli and behavioral success.

Neural circuitry is, of course, further modified in postnatal life according to individual experience by mechanisms of neural plasticity. Taking advantage of experience accumulated during childhood and adult life allows individuals to benefit from circumstances in innumerable ways that contend with the challenges in the world more successfully than would be possible using inherited circuitry alone. Changes in neural connectivity arising in phylogeny and ontogeny provide reasonably well-understood mechanisms (natural selection and neural plasticity) for instantiating neural circuitry on the basis of species and personal history.

Perceptions as Reflex Responses. If vision operates by associating retinal images with useful behaviors on the basis of past experience, then the mechanisms for perceptions and their genesis are best understood as a reflex response. Although the concept of a reflex is not precisely defined, it alludes to behaviors such as the knee-jerk response that depend on the "automatic" transfer of information through previously established circuitry. The advantages of reflex responses are clear: after natural selection and neural plasticity have done their work over evolutionary and individual time, the nervous system can respond with greater speed and accuracy.

It does not follow, however, that reflex responses are in any sense simple or that they are limited to "lower-order" neural circuitry. Sherrington, who worked out spinal reflex circuits in the first part of the 20th century, was well aware that the concept of a simple reflex is, in his terms, a "convenient . . . fiction" (31). As Sherrington was quick to point out, "all parts of the nervous system are connected together and no part of it is ever capable of

reaction without affecting and being affected by other parts" (31). If the framework offered here is right, then conceiving of vision in terms of automatic responses that are fully determined by existing circuitry whose function can nevertheless be influenced by many brain regions seems appropriate.

An objection to the idea that vision is reflexive might be that, because perceptions can be elicited by stimuli that have never been experienced, extant neural circuitry could not underwrite vision. However, every natural visual stimulus is literally unique. As long as a light stimulus is within the physical limits that humans can react to, a visual system that evolved empirically will rely on the sum total of previous experience with similar stimuli to generate a perception in accord with the relative success of past behavior.

In short, if visual system connectivity has been established on a wholly empirical basis by natural selection and activity-dependent plasticity, there is no reason to think that visual perceptions are anything more than reflexive responses determined by neural associations previously established according to the relative success of species and individual behavior.

Other Empirical Approaches. The influence of past experience on visual perception has played an important part in theories of vision at least since Helmholtz drew a distinction between sensation and perception, arguing that experience must resolve ambiguity in visual stimuli by generating "unconscious inferences," which lead to perceptions more likely to accord with reality (32). More recently, the role of past experience has played an important role in gestalt (33) and ecological (34) theories of perception, as well as in Bayesian models of vision (35–40).

Consider, for example, work predicated on Bayesian models. Bayes' theorem is used to assess the probability of an event, A, given another event or condition, B. Because this conditional probability depends on prior evidence about events A and B, it is referred to as the posterior probability. Thus, Bayes' theorem states that the posterior probability depends on the probability of B given A (called the likelihood) and critically, on the probability of A absent information about B (the prior) divided by the probability of B absent any information about A. This theorem is expressed as (Eq. **1**)

$$P(A \mid B) = \frac{P(B \mid A) \bullet P(A)}{P(B)}. \qquad \textbf{[1]}$$

In vision research, Bayes' theorem has been used to assess the most likely cause of an ambiguous or noisy retinal stimulus (35–40). Typically, A would stand for a physical state of the world, and B would be a retinal image. Assuming that representing the most probable state of the world in perception would benefit humans or other visual animals, a "Bayesian observer" would see the most likely physical source of an image, as specified by the highest value in a distribution of posterior probabilities of A, thus taking into account image ambiguity (i.e., that more than one source could have given rise to the image B).

If the theory that we have presented is right, however, the inverse problem prevents a Bayesian approach from explaining perception, much less speaking to the underlying organization of the visual system. Because the inverse problem precludes an assessment of physical properties of the world by the visual system (see Fig. 1), the relative probabilities of the states of the world as such are irrelevant to visual perception. For a Bayesian framework of vision to work, the relevant Bayesian priors would have to be useful behaviors, not states of the world. In theory, a wholly empirical strategy could be formulated in Bayesian terms; however, because the history of successful behaviors over evolutionary and individual time cannot be known, this exercise would have little use (*SI Text*).

Similar concerns apply to filter-based interpretations of vision. According to this view, the visual system evolved to extract as much information from the retinal image as possible, while minimizing the circuitry, signaling, and metabolic cost needed for efficient coding. Thus, rather than encoding the information at each point in the retinal image independently (which would result in redundant neural responses), in this conception the visual system filters images and extracts salient features like oriented edges according to the statistical characteristics of images arising from natural scenes at multiple spatial and temporal scales (19, 41–46). Although relying on filters to extract the salient features of images has provided much insight into how image information is likely to be related to receptive field properties and efficient information transfer, these useful strategies do not address the inverse problem and therefore, cannot explain visual perception.

Conclusion. The fundamental challenge for understanding biological vision is the inverse problem: light falling on the retina inevitably conflates the contributions to the stimulus arising from the physical features and conditions of real-world objects, thereby precluding the sources of visual stimuli from being determined. As a result, the significance of retinal images for visually guided behavior must be resolved empirically. The evidence briefly described here implies that, as a means of contending with the inverse problem, retinal stimuli trigger neuronal activity in circuits that have been fully determined by accumulated species and individual experience with the behavioral success (or failure) arising from interactions with the environment over time. The result is circuitry that reflexively generates perceptions based entirely on history. The evidence that the visual system operates in this counterintuitive way is the ability of data drawn from proxies for this history to explain what we actually see.

1. Palmer S (1999) *Vision Science: From Photons to Phenomenology* (MIT Press, Cambridge, MA).
2. Hubel DH, Wiesel TN (2005) *Brain and Visual Perception* (Oxford University Press, New York).
3. Purves D, Lotto B (2011) *Why We See What We Do Redux: A Wholly Empirical Theory of Vision* (Sinauer Associates, Sunderland, MA).
4. Gelb A (1929) Die "Farbenkonstanz" der Sehdinge. *Handbuch der normalen und pathologischen Physiologie*, eds von Bethe WA, von Bergmann G, Embden G, Ellinger A (Springer, Berlin), pp 594–678.
5. Gilchrist AL (2009) *Seeing in Black and White* (Oxford University Press, Oxford).
6. Purves D, Lotto B (2003) *Why We See What We Do: An Empirical Theory of Vision* (Sinauer, Sunderland, MA).
7. Yang Z, Purves D (2004) The statistical structure of natural light patterns determines perceived light intensity. *Proc Natl Acad Sci USA* 101:8745–8750.
8. Howe CQ, Purves D (2005) *Perceiving Geometry: Geometrical Illusions Explained by Natural Scene Statistics* (Springer, New York).
9. Robinson JO (1998) *The Psychology of Visual Illusions* (Dover, New York).
10. Wundt W (1961) Contributions to the theory of sensory perception. *Classics in Psychology*, ed Shipley T (Philosophical Library, New York), pp 51–78.
11. Shipley WC, Nann BM, Penfield MJ (1949) The apparent length of tilted lines. *J Exp Psychol* 39:548–551.
12. Pollock WT, Chapanis A (1952) The apparent length of a line as a function of its inclination. *Q J Exp Psychol* 4:170–178.
13. Cormack EO, Cormack RH (1974) Stimulus configuration and line orientation in the horizontal-vertical illusion. *Percept Psychophys* 16:208–212.
14. Craven BJ (1993) Orientation dependence of human line-length judgements matches statistical structure in real-world scenes. *Proc Biol Sci* 253:101–106.
15. Howe CQ, Purves D (2002) Range image statistics can explain the anomalous perception of length. *Proc Natl Acad Sci USA* 99:13184–13188.
16. Howe CQ, Purves D (2005) Natural-scene geometry predicts the perception of angles and line orientation. *Proc Natl Acad Sci USA* 102:1228–1233.
17. Marr D, Ullman S (1981) Directional selectivity and its use in early visual processing. *Proc R Soc Lond B Biol Sci* 211:151–180.
18. Hildreth E (1984) *The Measurement of Visual Motion* (MIT Press, Cambridge, MA).

19. Adelson EH, Bergen JR (1985) Spatiotemporal energy models for the perception of motion. *J Opt Soc Am A* 2:284–299.
20. Movshon JA, Adelson EH, Gizzi MS, Newsome WT (1986) The analysis of moving visual patterns. *Pattern Recognition Mechanisms*, eds Chagas C, Gattass R, Gross C (Springer, New York), pp 148–163.
21. Ullman S, Yuille AL (1989) Rigidity and the smoothness of motion. *Image Understanding*, eds Ullman S, Richards W (Alex Publishing Corporation, Norwood, NJ).
22. Nijhawan R (1994) Motion extrapolation in catching. *Nature* 370:256–257.
23. Whitney D, Murakami I (1998) Latency difference, not spatial extrapolation. *Nat Neurosci* 1:656–657.
24. Yuille AL, Grzywacz NM (1998) A theoretical framework for visual motion. *High-Level Motion Processing: Computational, Neurobiological, and Psychophysical Perspectives*, ed Watanabe T (MIT Press, Cambridge, MA), pp 187–211.
25. Eagleman DM, Sejnowski TJ (2000) Motion integration and postdiction in visual awareness. *Science* 287:2036–2038.
26. Eagleman DM, Sejnowski TJ (2002) Untangling spatial from temporal illusions. *Trends Neurosci* 25:293.
27. Wuerger S, Shapley R, Rubin N (1996) On the visually perceived direction of motion by Hans Wallach: 60 years later. *Perception* 25:1317–1367.
28. Wojtach WT, Sung K, Truong S, Purves D (2008) An empirical explanation of the flash-lag effect. *Proc Natl Acad Sci USA* 105:16338–16343.
29. Wojtach WT, Sung K, Purves D (2009) An empirical explanation of the speed-distance effect. *PLoS ONE* 4:e6771.
30. Sung K, Wojtach WT, Purves D (2009) An empirical explanation of aperture effects. *Proc Natl Acad Sci USA* 106:298–303.
31. Sherrington SC (1947) *The Integrative Action of the Nervous System* (Yale University Press, New Haven, CT), 2nd Ed.
32. Helmholtz HLFv (1924) *Helmholtz's Treatise on Physiological Optics* (The Optical Society of America, Rochester, NY), 3rd Ed.
33. Kohler W (1947) *Gestalt Psychology: An Introduction to New Concepts in Modern Psychology* (Liveright, New York).
34. Gibson JJ (1979) *The Ecological Approach to Visual Perception* (Lawrence Erlbaum, Hillsdale, NJ).
35. Knill DC, Richards W (1996) *Perception as Bayesian Inference* (Cambridge University Press, Cambridge, UK).
36. Geisler WS, Kersten DC (2002) Illusions, perception and Bayes. *Nat Neurosci* 5:508–510.
37. Rao RPN, Olshausen BA, Lewicki MS (2002) *Probabilistic Models of the Brain: Perception and Neural Function* (MIT Press, Cambridge, MA).
38. Weiss Y, Simoncelli EP, Adelson EH (2002) Motion illusions as optimal percepts. *Nat Neurosci* 5:598–604.
39. Kersten D, Yuille A (2003) Bayesian models of object perception. *Curr Opin Neurobiol* 13:1–9.
40. Knill DC, Pouget A (2004) The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends Neurosci* 27:712–719.
41. van Santen JP, Sperling G (1984) Temporal covariance model of human motion perception. *J Opt Soc Am A* 1:451–473.
42. Watson AB, Ahumada AJ, Jr. (1985) Model of human visual-motion sensing. *J Opt Soc Am A* 2:322–341.
43. Field DJ (1994) What is the goal of sensory coding? *Neural Comput* 6:559–601.
44. Olshausen BA, Field DJ (1997) Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Res* 37:3311–3325.
45. Olshausen BA, Field DJ (2000) Vision and the coding of natural images. *Am Sci* 88:238–245.
46. Basole A, White LE, Fitzpatrick D (2003) Mapping multiple features in the population response of visual cortex. *Nature* 423:986–990.