Neurological models of size scaling

Helen E. Ross
Department of Psychology, University of Stirling, Stirling FK9 4LA, United Kingdom. **h.e.ross@stir.ac.uk**
http://www.stir.ac.uk/psychology/staff/her1

Abstract: Lehar argues that a simple Neuron Doctrine cannot explain perceptual phenomena such as size constancy but he fails to discuss existing, more complex neurological models. Size models that rely purely on scaling for distance are sparse, but several models are also concerned with other aspects of size perception such as geometrical illusions, relative size, adaptation, perceptual learning, and size discrimination.

Lehar argues (sect. 2.2 and elsewhere) that there are no adequate neurological models to explain why we see the world the way we do, and that theorists have ignored the discrepancies between the proximal stimulus and our perceptual experience. He then presents a computational model to describe our perceptual experience of hyperbolic space. He rightly complains about the shortage of neurological models for size and shape constancy but he fails to discuss the models that do exist.

Psychologists have long been interested in size scaling, or discrepancies between perceived size and image size: The phenomena include size constancy, geometrical illusions, optical distortions, adaptation, and aftereffects. The classical account of size constancy maintains that size is scaled for distance in a quasi-geometric manner (the size-distance invariance hypothesis); this account is not productive of neurological models because it assumes that retinal image size is "correctly" encoded in the visual cortex and that the image is then scaled for distance in some unexplained "cognitive" manner. Kirschfeld (1999) argues that the image representation has to be scaled for distance neurologically before it enters consciousness and that this might be done in area V4. He notes that Dobbins et al. (1998) found that some neurons in this area varied their response to the angular size of lines depending on viewing distance.

The idea that image size is transformed at some preconscious stage of visual processing by mechanisms other than distance scaling (e.g., McCready 1985) may be more fruitful. Stuart et al. (1993) proposed a computational model based on broadly tuned layers of size detectors, which could account both for Weber's law in size discrimination and for the biasing effects of geometrical illusions; however, they did not extend the model to include scaling for distance. The main alternative approach to size constancy – generally supported by Gibsonians – is that object sizes are scaled in relation to the surrounding spatial scale. This approach has the advantage of embracing other size illusions in addition to size constancy and it is more productive of neurological models. Size contrast illusions have been attributed to adaptation of cells that detect spatial frequency or to other neural interactions in the brain (see Gillam 1998). However, spatial frequency is not the same thing as image size (the distance across an image), so spatial frequency models are unhelpful for general models of size perception.

Andrews (1964) proposed a perceptual learning model of size calibration in which the brain corrects the metric of the visual field according to the most recent information and attempts to equalize the spacing of contours. This would allow for learning in addition to explaining some illusions, aftereffects, and size constancy. Richards (1977) suggested that simple cells in the cortex might respond to relative rather than absolute size and he also discussed the properties necessary for the neural basis of size constancy.

Some authors have attempted to explain size constancy through the enlargement of perceived size for the central part of the visual scene, which occurs because the representation of the central part of the retinal image covers more cortical cells at later stages of analysis. Such an idea is based on the anatomical fact of cortical magnification, which enhances acuity for central vision. The fovea contains more densely packed cone cells than the surrounding area and it projects to a relatively larger region of the primary visual cortex. Schwartz (1980) incorporated this idea into his model of size constancy. When an observer fixates a distant object, it forms a small image in central vision, whereas close objects form larger images that spread further into the periphery: The small central image is therefore expanded neurologically relatively more than the larger image. Such a mechanism might contribute marginally to size constancy, but it fails to explain how objects of the same angular size can appear different in size even when both are viewed in central vision.

An example of this problem is the moon illusion (see Ross & Plug 2002). The moon illusion is the apparent enlargement of the sun or moon when low on the horizon compared with its size when higher in the sky; the effect is similar to size constancy but is hard to explain by the usual "scaling for distance" account. The difficulty is that the low moon appears nearer than the high moon, whereas size-distance invariance requires it to appear further. Trehub (1991, pp. 242–47) developed the "retinoid" model, which could account for both size constancy and the moon illusion. He argued that size magnification is expensive in neurological terms because it involves the use of more networks of cells. The brain husbands its resources by magnifying only the most "ecologically relevant" parts of the scene – that is, objects in the near distance when looking horizontally, and close overhead when looking up. Humans cannot normally interact with celestial objects or with distant terrestrial objects, so the images for such objects can safely be left relatively small. Size constancy is therefore poor for far horizontal distances and even poorer when looking upwards. The three-dimensional representation of distance is also shrunk vertically in comparison with horizontally, again for the purpose of minimizing neural resources. Distance is computed within the three-dimensional retinoid system and is represented by "sheets" of cells; the extent of size magnification is linked to the distance plane onto which the image is mapped. This biased mapping of the visual scene onto brain structures is largely the result of human evolution, but it can be further modified by individual experience.

There are neuropsychological findings that support multiple representations of three-dimensional space (see Previc 1998). There are also findings on micropsia and hemineglect that give clues as to how and where size might be coded (see Kassubek et al. 1999). Lehar may be correct that a simple Neuron Doctrine cannot account for size scaling, but more complex neurological models show promise.

Spatial phenomenology requires potential illumination

James A. Schirillo
Department of Psychology, Wake Forest University, Winston-Salem, NC 27109. **schirija@wfu.edu**

Abstract: Collapsing three-dimensional space into two violates Lehar's "volumetric mapping" constraint and can cause the visual system to construct illusory transparent regions to replace voxels that would have contained illumination. This may underlie why color constancy is worse in two dimensions, and argues for Lehar to revise his phenomenal spatial model by putting "potential illumination" in *empty* space.

Lehar's phenomenological description of space neglects the fact that *empty* space is actually *full* of illumination. For example, if a cast shadow crosses half of this page and you move your finger from a word under shadow to one under full illumination, you are not surprised when your finger crosses the shadow, even though your finger is closer in depth than the page. This is because every voxel between your eye and the page contains some amount of light. It is unfortunate that Lehar overlooks this fact, because he correctly asserts that depth information is volumetric, whereas current neurological models fail to "represent transparency[,] with multiple depth values at every single (x,y) location, or to represent the experience of empty space between the observer and a visible object" (sect. 3, para. 5). These same models also ignore that every voxel of "empty" space contains light of some intensity and chromaticity.

This confusion probably results from naïvely accepting the popular notion that humans care only about the location and qualities of objects, making the perception of illumination irrelevant. This assumption is so prevalent that much of color research is devoted to determining how the visual system "discounts the illuminant." However, a viable solution to the Gestalt problem of color constancy will emerge only with a more complete description and understanding of how we subjectively *experience* illumination. Ironically, Lehar's aspiration to describe the subjective experience of spatial vision in terms comparable to those of color vision reveals that current color vision research is also in peril. That is, he claims that color phenomena are reducible to hue, intensity, and saturation because that is how the brain represents them physiologically (sect. 2.3). Yet models of hue, intensity, and saturation cannot be the "primitives of raw conscious experience" (sect. 4, para. 3), in that these qualities remain invariant as illumination changes across space.

This confound is apparent when Lehar discusses his Figure 1 as containing "explicit volumes, bounded by colored surfaces, embedded in a spatial void" (sect. 5.1, para. 2), where "every point can encode either the experience of transparency or the experience of a perceived color at that location" (sect. 6). His more accurate intuition is that there are also intermediate states between transparent and opaque "to account for the perception of semitransparent surfaces" (sect. 8.1, para. 1). I suggest that Lehar consider filling these semitransparent voxels with "potential illumination" "at multiple depth values at every single (x,y) location." This would also strengthen his second and third conclusions that "volumes of empty space are perceived with the same geometrical fidelity as volumes of solid matter" and that "multiple transparent surfaces can be perceived simultaneously" (sect. 10, points 2 and 3). Having semitransparent voxels contain "potential illumination" is a more parsimonious description of the void between your eyes and this page. You can actualize the "potential illumination" of these voxels by placing your finger in front of any shadow cast on the page. More accurately, Lehar's phenomenological model allows *only* the plane of voxels directly in front of a given surface to contain cast shadows (i.e., less illumination), because the voxels that compose the surface must be the color of the opaque surface itself (sects. 5.1, 8.1).

Note that this concept is not merely peripatetic (Aristotle 1976) or an ether explanation, in that we are always subjectively aware of the illuminant. For example, by looking from your illuminated reading room into a dark hallway, your subjective experience is not only that the hallway walls are under less illumination but also that the space itself contains less light. In this way, "potential illumination" can also address why color constancy differs in two- versus three-dimensional scenes. For example, Gilchrist (1977) had observers look through a pinhole into a room containing a doorway into a second room. The near room was dimly illuminated and the far room was highly illuminated. Attached to the door frame were several papers, arranged so that a mid-gray paper appeared either adjacent to the door frame or (with its corners removed) on the far room's back wall. The lightness of the paper shifted in the direction of lightness constancy depending on whether it appeared on the door frame or on the far wall. Schirillo et al. (1990) generated equivalent stimuli in two dimensions using a stereoscopic cathode ray tube (CRT) and stereoscope, yet this replication produced only a fraction of Gilchrist's constancy. I hypothesize that this occurred because stereoscopic space does not contain the actual voxels of either high (e.g., near room) or low (e.g., far room) illumination. In essence, Schirillo and colleagues failed to preserve Lehar's necessary condition of "volumetric mapping" (target article, Fig. 1D).

The ubiquitous use of CRT images reduces scenes to Alberti's window, which retains perspective cues but eliminates Lehar's requirement that space be volumetric. This obfuscates the color constancy paradox, in that these voxels contain illumination. For example, Adelson's (1993) famous wall-of-blocks illusion contains cubes of identical luminance that appear dissimilar in lightness and concomitantly under illusory transparent stripes. Logvinenko et al. (2002) eliminated both the appearance of transparency and the lightness illusion by constructing a three-dimensional version of Adelson's two-dimensional display. I hypothesize that the visual system does not add a transparent veil to Logvinenko's display because it already ascribes illumination to every voxel in space. However, when Adelson eliminates such volumes but retains the same spatial geometry via X-junctions, the visual system reconstructs the volume to contain regions of illusory transparency (i.e., illumination). Consequently, Lehar's improved spatial model requires a phenomenal description of *empty* space that includes "potentially illuminated" voxels.

If vision is "veridical hallucination," what keeps it veridical?

Peter Ulric Tse

Department of Psychological and Brain Sciences, Dartmouth College, Hanover, NH 03755. **Peter.Tse@dartmouth.edu**
http://www.dartmouth.edu/artsci/psych/faculty/tse.html

Abstract: If perception is constructed, what keeps perception from becoming mere hallucination unlinked to world events? The visual system has evolved two strategies to anchor itself and correct its errors. One involves completing missing information on the basis of knowledge about what most likely exists in the scene. For example, the visual system fills in information only in cases where it might be responsible for the data loss. The other strategy involves exploiting the physical stability of the environment as a reference frame with respect to which the eyes and body can move.

> *[S]pace and time are only forms of sensible intuition, and hence are only conditions of the existence of things as phenomena . . . we can have no cognition of an object, as a thing in itself, but only as an object of sensible intuition, that is, as phenomenon*
> —Immanuel Kant (*Critique of Pure Reason*, 1781)

Lehar develops the Kantian insight that perception is (1) entirely a mental construction; (2) lacks access to the world-in-itself to determine the accuracy of its representations; and (3) is only possible given an internal framework of space-time that permits sensory input to be interpreted as occurring in an external space-time. Here I focus on how the brain can construct true information about the world when there is no way to judge objectively whether that information is true by comparing that information to the world-in-itself.

To create veridical information, the visual system must compensate for errors, data loss, and processing bottlenecks imposed by its imperfect design. It has nothing but the ambiguous, incomplete, and noisy image to determine whether it has made an error. It must therefore know what types of image cues indicate errors and it must have strategies to correct errors. The visual system corrects itself only when it is responsible for errors or data loss. It compensates for its own likely errors using two strategies. One is to rely on world knowledge, and the other is to assume that the world is stable.

For example, when does the visual system fill in missing phenomenal features and when does it merely note that completion takes place without filling-in (see Fig. 1)? Filling-in occurs when the information that is missing from the image is missing because of the visual system's own failure to detect it. The visual system follows the principle "no news isn't necessarily bad news when there was no way to get the news in the first place." The visual system functions as if it knows that it does not always have adequate in-