If Lehar intends to create a Gestalt-oriented theory of perception, he has to have in mind that according to the classics of Gestalt theory, the phenomenological Gestalten are the consequences of both internal (neural) and external (stimulus) constraints (Koffka 1935; Köhler 1920/1938; 1927/1971; 1947). Simply speaking, the perceptual system tends to attain the maximum efficiency with the minimum investment (internal neural economy), but the minima and maxima will always be relative to the given stimulus conditions (external stimulus organization). The effect of external "control" of a perceptual economy is an articulation of more or less *prägnant Gestalten,* or as Wertheimer stated in his famous Law of Prägnanz, the phenomenal organization of a percept will be as "good" as the prevailing conditions allow (cf. Koffka 1935).

Bursting the bubble: Do we need true Gestalt isomorphism?

Niall P. McLoughlin

Department of Optometry and Neuroscience, University of Manchester, Institute of Science and Technology (UMIST), Manchester M60 1QD, United Kingdom. **n.mcloughlin@umist.ac.uk**

Abstract: Lehar proposes an interesting theory of visual perception based on an explicit three-dimensional representation of the world existing in the observer's head. However, if we apply Occam's razor to this proposal, it is possible to contemplate far simpler representations of the world. Such representations have the advantage that they agree with findings in modern neuroscience.

Lehar proposes to model visual perception using his subjective visual experience as his source of data. He proposes a perceptual modeling approach because "conventional concepts of neural processing offer no explanation for the holistic global aspects of perception identified by Gestalt theory" (target article, Abstract). This allows him conveniently to ignore current research in visual neuroscience while concentrating on the central issues of the representation of the visual field and of our subjective visual experiences. As he correctly points out, the world we see and experience surrounding us exists only as nerve impulses within our head. Lehar proposes that because our subjective experience of the world is that of a high-resolution three-dimensional volume, and because this representation must exist in our heads, it must therefore be some form of a high-resolution three-dimensional structure. However, this does not necessarily follow. For example, on a computer system it is possible to generate a sparse representation of the world into which it is placed so that the computer could interact with objects in the world in a meaningful manner. Objects could be represented as tokens at such-and-such x, y, and z location, and so forth. There would be no explicit representation of empty space within this sparse representation. Who is to say what the subjective experience of the computer might be?

There is no doubt that my subjective experience of the world is that of a three-dimensional solid environment which I perceive in equal detail in all directions. Yet, as visual scientists and practiced observers, we know that this is patently not the case. Each of our eyes responds to incoming photons in a non-uniform manner and this non-uniformity is further exaggerated in the cortex. The over-representation of the fovea is magnified between the retina and cortex, and the multiple interconnected cortical regions amplify this distinction even further. Most naïve observers are surprised to discover that they have a fovea and amazed that they have a blind spot in each eye. How do we fool ourselves?

The very fact that we are genuinely fooled (until we make careful observations) calls into question the use of subjective experience as the basis for theories of visual perception. Furthermore, although the Neuron Doctrine is indeed the foundation for most modern neuroscience research, I refute the notion that this doctrine implies purely feed-forward models of neurocomputation. Certainly, recent findings in both neuroanatomy (e.g., Angelucci et al. 2002; Bosking et al. 1997) and neurophysiology (Kapadia et al. 2000; Levitt & Lund 1997) emphasize the roles played by feedback and lateral connections in visual processing. Likewise, a number of popular modern computational theories make use of feed-forward, feedback, and lateral connections (e.g., Grossberg 1994). If a Gestalt Bubble model subserves perception, then why do we have so many visual areas, each containing a retinotopic map of visual space?

Is there any evidence for Gestalt-like processes at work neurophysiologically? Recent electrophysiological recordings from as early as the lateral geniculate and V1 have found interactions well outside the classical receptive field (e.g., Blakemore & Tobin 1972; Felisberti & Derrington 2001; Jones et al. 2000; 2001; Kapadia et al. 2000; Levitt & Lund 1997; Solomon et al. 2002; Stettler et al. 2002). Although the source of these interactions (whether they are mediated by feedback or by lateral connections) remains to be elucidated, it is clear that many aspects of grouping, completion, and emergence may well arise from such nonlocal interactions. In addition, recent neurophysiological studies in the primate (e.g., Livingstone & Hubel 1988) suggest that different aspects of a visual scene are represented primarily in different visual streams and areas. Although there is some disagreement as to the amount of segregation of function, numerous neuropsychological studies in humans back up the suggestion that multiple representations exist for different attributes and/or functional roles. One such patient studied by Humphrey and Goodale (1998) suffered from visual-form agnosia (Farah 1990). She was unable to discriminate between visual forms, let alone recognize her friends and family, yet her color vision was close to normal and she could recognize shapes when placed in her hands. Such case studies suggest that the brain encodes the external world using multiple representations, each one perhaps subserving a different role or task rather than a single isomorphistic one.

What Lehar seems to have forgotten is that the high-resolution representation is generated only when we pay attention to the input and focus our eyes on the object or texture under inspection. We need not represent even our immediate environment in high resolution unless we need to interact directly with it. Why waste time and space representing the world in vivid detail when we interact with only a small part of it at any one time? Surely our central representations should be goal-directed. We can always direct our vision to different locations in a scene to find out what is there, and given that most useful scenes are dynamic, why waste effort representing space in high resolution when it is constantly changing? O'Regan (1992) argued along a similar line when he suggested that "seeing constitutes an active process of probing the environment as though it were a continuously available *external memory*" (p. 484, emphasis in original). He suggests that seeing does not involve the reification of a three-dimensional spatial representation of the external world in the observer's head but rather depends on one's ability to interrogate the environment through directed eye movements. It may well be that we have a fuzzy three-dimensional representation of the external world in our heads that we use to help direct eye movements, but I remain to be convinced that we would need or want anything more complex. If we need the detail, we look.

Given the lack of physiological evidence for such a complex and computationally expensive representation, coupled with the lack of necessity for such a complete representation, Occam's razor suggests we burst this Gestalt Bubble model.