Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Physics of Life Reviews ••• (••••) •••–•••
PHYSICS of LIFE reviews
www.elsevier.com/locate/plrev

Comment

# Adding Gestalt to the picture
# Comment on “Move me, astonish me… delight my eyes and brain: The Vienna Integrated Model of top-down and bottom-up processes in Art Perception (VIMAP) and corresponding affective, evaluative, and neurophysiological correlates” by Matthew Pelowski et al.

Johan Wagemans

*Laboratory of Experimental Psychology, Department of Brain & Cognition, University of Leuven (KU Leuven), Belgium*

Received 1 June 2017; accepted 2 June 2017

Communicated by L. Perlovsky

*Keywords:* Art perception; Gestalt; Aesthetics; Visual processing; Dynamics; Phenomenology

Matthew Pelowski and his colleagues from the Helmut Leder lab [17] have made a remarkable contribution to the field of art perception by reviewing the extensive and varied literature (+300 references) on all the factors involved, from a coherent, synthetic perspective—The Vienna Integrated Model of top-down and bottom-up processes in Art Perception (VIMAP). VIMAP builds on earlier attempts from the same group to provide a comprehensive theoretical framework, but it is much wider in scope and richer in the number of levels and topics covered under its umbrella. It is particularly strong in its discussion of the different psychological processes that lead to a wide range of possible responses to art—from mundane, superficial reactions to more profound responses characterized as moving, disturbing, and transformative. By including physiological, emotional, and evaluative factors, the model is able to address truly unique, even intimate responses to art such as awe, chills, thrills, and the experience of the sublime. The unique way in which this rich set of possible responses to art is achieved is through a series of five mandatory consecutive processing steps (each with their own typical duration), followed by two conditional additional steps (which take more time). Three processing checks along this cascade lead to three more or less spontaneous outcomes (<60 sec) and two more time-consuming ones (see their Fig. 1 for an excellent overview). I have no doubt that VIMAP will inspire a whole generation of scientists investigating perception and appreciation of art, testing specific hypotheses derived from this framework for decades to come.

In spite of my high appreciation for this wonderful achievement, I see two major problems in the current framework.

The first specifically concerns the processing of the formal aspects of the artwork, explicitly labeled as low-level artwork-derived features, although “these might include our responses to symmetry, lines, colors, balance, or other

DOI of original article: http://dx.doi.org/10.1016/j.plrev.2017.02.003.
*E-mail address:* johan.wagemans@kuleuven.be.

http://dx.doi.org/10.1016/j.plrev.2017.06.009

factors which emerge when we first look at an artwork, as well as their combination into more complex scenes and patterns." (p. 7) These formal aspects contain a mixed bag of features—some that are definitely established in low-level visual areas (e.g., lines, colors) and others that are much more likely to require integration at somewhat higher levels (e.g., patterns, symmetry, balance). Nevertheless, "they are often connected to automatic, bottom-up processing (...) below the level of awareness." (p. 7) I think it is mistaken to assume that all these formal aspects of artworks are established quickly and automatically (even outside of awareness), and that the processes involved are exclusively implemented in low-level mechanisms. When considering formal aspects of artworks, it is useful to make a distinction between global image properties such as homogeneity or heterogeneity of edge orientations, peaks in the color histogram, self-similarity, 1/F structure, and so forth (e.g., [8,20,19]) and image properties such as symmetry and balance, which require explicitly encoded spatial relationships between (sets of) localized features, patches, blobs, shapes, and proto-objects (for recent reviews, see [3,23]). While the former can be computed on the distribution of the outputs of low-level filters, the latter are established only in intermediate or higher areas (e.g., V4, the Lateral Occipital Cortex).

Admittedly, in their discussion of Stage 3 (implicit memory integration), the authors acknowledge that this stage "involves a more focused period of attention, in which low-level features are segregated or grouped to form larger" (p. 16) "units corresponding to intermediate stages of vision" (p. 17) and they suggest that some higher-order characteristics like structural unity, order, entropy, or Prägnanz, although potentially first detected in Stage 2, could be further processed in Stage 3 before they lead to positive appraisal and affect. This idea of reiterative processing of image properties at multiple levels in the cortical hierarchy, during multiple sweeps or cycles, appears to be essential to explain how the same image could yield such varied responses in the same viewer at multiple instances of viewing the same image or in different viewers with their own perceptual styles, skills or art experience. The most interesting artworks usually require somewhat more extensive processing to grasp the overall composition and to appreciate higher-order properties like the balance between different parts of a scene or the dynamic switching of multiple local figure-ground organizations [2,14,15]. Experienced viewers often exploit extensive looking, moving closer or further away from a painting, alternating blurred and sharp vision, etc. to enable multiple structures and organizations to emerge. This clearly does not match the simple labels of 'automatic', 'bottom-up' and 'low-level' assigned to formal aspects of artworks.

The second, more general and therefore more serious problem is that VIMAP is founded on a simplified, somewhat outdated view on the distinction between "bottom-up processing of artwork features (form, attractiveness) and top-down contributions of memory, personality, and context" (p. 4). Moreover, the whole processing account is rather strictly sequential and feedforward. For instance, "Stages 1–4 would" (p. 40) "concern primarily formal artwork features, which are automatically processed and would feed into the cognitive mastery stage" (p. 41). This contrasts with more modern views on the dynamics of processing throughout the cortical hierarchy. One of these is the so-called Reverse Hierarchy Theory [9], which dissociates early implicit from later explicit perception. In particular, the theory proposes that visual processing goes through a fast feedforward sweep of processing, yielding rapid awareness of the conceptual gist of the scene at the highest level in the hierarchy. Following this first implicit perception ('vision at a glance'), feedback connections to low-level visual areas allow more refined, recurrent processing of local details ('vision with scrutiny'). This slower processing occurs in low-level visual areas, where the neural maps have retinotopy and neurons have sufficiently small receptive fields to capture the details. So, Reverse Hierarchy Theory also dissociates the temporal early-late distinction from the structural distinction between low- and high-level areas in the brain, and it focuses on dynamic iterations between bottom-up and top-down processes ('re-entrant processing').

Another, even more radical view on the dynamics of cortical processing is offered by the predictive coding framework [5,7]. Predictive coding assumes that the brain builds a so-called generative model about the environmental causes of the perceptual inputs it receives. It infers these causes by making a best guess (or 'prediction') about incoming inputs at each point in time and checking whether the predicted sensory activity corresponds with that actually received through the senses. If not, it will attempt to reduce this mismatch (or 'prediction error') by adjusting its prediction about the state of the environment and adapting its generative model for the current context accordingly. Within this scheme, these models are hierarchically structured [24], where higher levels are capable of capturing patterns in perceptual inputs that have larger spatial or temporal spans. In this view, processing does not start from the input but from the brain's predictions. By assuming that an active reduction of prediction errors is experienced as pleasurable, this predictive coding framework can explain the dynamics of aesthetic pleasures in the perception of even static artworks [22]. Moreover, this prediction-error account of visual art appreciation (see also [4,6,10,16]) can

Please cite this article in press as: Wagemans J. Adding Gestalt to the picture. Phys Life Rev (2017), http://dx.doi.org/10.1016/j.plrev.2017.06.009