Available online at www.sciencedirect.com

SCIENCE DIRECT®

Vision Research 45 (2005) 381–385

Vision Research

www.elsevier.com/locate/visres

Letter to the editor

Object substitution and its relation to other forms of visual masking: reply to James Enns

In a recent paper (Enns, 2004), James Enns demonstrated nearly equivalent and strong backward-masking effects of various types of visual masks at longer target and mask temporal intervals (e.g., 150 ms), including strong masking with four dots considered to be a weak mask in standard masking conditions. The principal requirement for strong masking at long temporal intervals was that target had to be included among several other visual objects, which guaranteed that spatial attention was dispersed over an extended visual region in space. If time intervals were short (e.g., 50 ms), allowing integration of target and mask features, different masks had different effects. If attention was pre-focused on the spatial location of the target by precues, masking was negligible. James Enns (2004) claims that these findings pointed to a new understanding of masking based on the separate processes of object formation and object substitution. As a first item of my present reply, I will argue that the claim for novelty is clearly overstated.

In addition, Enns (2004) listed four features of masking as especially difficult to explain within standard theories of visual masking, the standard ones being (1) visual integration and perceptual confusion theories, (2) theories of processing interruption, (3) masking by competitive neural interaction. The four features difficult to explain by these theories being (i) no necessity of local contour interaction between target and mask for masking to occur, (ii) strong modulation of masking by a weak mask (e.g., four dots surrounding the target) by spatial attention, (iii) increase in the strength of masking with temporal extension of mask exposure beyond the offset of target, (iv) importance of the status of target and mask as individual objects for the expression of masking. James Enns claims that the only theory capable of explaining these features is the theory of substitution (Enns, 2004; DiLollo, Enns, & Rensink, 2000; Enns & DiLollo, 1997). As a second item of my reply, I will argue that Enns (2004) has overlooked one theory that should be considered in explaining why the visual system has preference for the sensory input that arrives relatively late in time regardless of the type of mask and why backward masking is sensitive to attentional manipulations.

In Enns (2004, Abstract, p. 1321) it is stated that "Results... pointed to a new understanding of masking based on the separate processes of *object formation* and *object substitution*". Also, his analysis "... provides strong support for the idea, derived from the object substitution theory of masking (DiLollo et al., 2000), that there are at least two distinct visual masking processes (Enns, 2004, p. 1328)". We are left with the view that this theoretical explanation is an invention from late 1990s and early 2000. Is it so?

Actually, already Turvey (1973), DiLollo, Lowe, & Scott (1974) and Hogben & DiLollo (1974) have described a picture of internal visual processing similar to the one promoted recently by Enns and his associates. Thus, in Hogben & DiLollo (1974) the two stages were termed as synthesis of forms and segregation of forms (implicitly suggesting the possibility of between-object competition). In Bachmann & Allik (1976), the formation stage followed by substitution stage was made more explicit. In that study, two spatially overlapping visual forms were presented with varying stimulus onset asynchronies. Subjects had to identify both forms. With very short SOAs both forms had approximately equal chance to be correctly perceived. At intermediate SOAs below 100 ms the first form perception dropped to a chance level while the second form dominated in explicit perception at a high level of identification rate. With long SOAs over 150 ms both forms were successfully identified. The identification function for the first form was U-shaped, reminiscent of typical metacontrast functions. If subjects attentively searched for the presence of pre-designated target objects within the pairs of forms, the U-shaped function for the first form disappeared and both forms were perceived at about equal level of correct responses. The theoretical explanation

0042-6989/$ - see front matter
doi:10.1016/j.visres.2004.08.011

suggested in that article (Bachmann & Allik, 1976) was essentially the same as it would be in any theory proposing feature integration and object replacement as the two principal processes underlying masking in visual processing. To prove the point, let me quote (Bachmann & Allik, 1976, p. 92): "...divergent feature analysis is carried out by feature detectors... The stimulus is analysed with regard to its orientation, location, size, frequency, length, and the other characteristics, ... All these elements begin to integrate. ... This triggers a concrete representation on the next, spatial objects', or iconic level. ... Where the pre-iconic stage is preattentive, the post-iconic operations are focal." And continuing from the next page (Bachmann & Allik, 1976, p. 93): "Suppose now that two stimuli, S1 and S2 enter in rapid succession into this analysing system. ... When the information on S2 reaches features level before the processing of S1 at this level is completed, then the features of both stimuli can be analysed in parallel. ... As a consequence of this simultaneous integration of features, a common iconic representation is formed. If the construction of both stimuli or their energy ratio is such as to permit read-out from the integrated image, then both stimuli have high probabilities of being recognised. ... When S1 is already represented at the iconic level, ... it will be categorised or encoded. ... When the features of S2 are integrated before the encoding of S1 is completed, then the succeeding item replaces the 'old' icon with the representation of a new object. ..., the subject must at the first opportunity... name 'triangle' to the disc which he sees. But this inconsistent outcome is ruled out by the internal consistency of brain functioning... The subject is unable to pay conscious attention to two objects at once, although they are represented at different levels. On the neurophysiological plane this is possibly done... by distortion of intercortical excitatory feedback loops..." In essence, this explanation suggests also a formation stage and a between-object competition stage that allows substitution of the initial object processing by the succeeding object processing. Thus what Enns terms as "new" maybe a rephrasing of the old.

My second item suggests that the theory of perceptual retouch (see Bachmann, 1984, 1994; Breitmeyer & Ögmen, 2000) should be also considered if we want to see whether object substitution theory (Enns, 2004; Di-Lollo et al., 2000; Enns & DiLollo, 1997) might be the only one capable of explaining the four features outlined in the beginning of the paper by Enns (2004) and capable of explaining his experimental results. Let me describe the foundations of the retouch theory before applying it for explaining backward masking and the four features of it claimed by Enns (2004) to be unexplainable by all other theories except substitution theory.

There is a widely accepted distinction between two brain systems: the sensory systems for stimulus-specific *content* and the systems for providing sufficient level (and frequency-pattern?) of cortical activation that is necessary for permitting a particular content to become explicitly represented (become conscious). The latter provides an *enabling factor* that is required for awareness, but does not directly contain specific contents of conscious experiences (Baars, 1995, 1997; Bachmann, 1984; Bogen, 1995; Llinás & Ribary, 2001; Rees, Kreiman, & Koch, 2002). The neurons of the content-specific system are termed "drivers" and the neurons of the conscious state systems belong to the class of "modulators" (Crick & Koch, 1998; Sherman & Guillery, 1998). Drivers that have small receptive fields and that respond to spatially localised stimuli with very short delays are modulated by the facilitating input from the content-free "modulators" of the so-called non-specific thalamus (Bachmann, 1984, 1994; Crick, 1984; Magoun, 1958; Purpura, 1970; Steriade, 1996a, 1996b; Steriade, Jones, & Llinás, 1990; Steriade, Jones, & McCormick, 1997). Drivers encode specific stimulus features such as size, orientation, color, motion, etc. The thalamic structures termed "non-specific" (e.g., intralaminar nuclei, pulvinar, nucleus reticularis, etc.) do not participate directly in the operations of encoding of the contents of specific sensory information. Although their efferent pathways are projected presynaptically onto specific cortical driver-neurons. Non-specific units modulate the level of activity of the drivers, no matter what were the specific signals that evoked the activity within the specific system in the first place. Therefore the signal-to-noise ratio of the activity of cortical driver-units that signal the presence of some sensory feature (or combination of features) is altered. Mostly, this is done through the excitatory synapses and the effects include increase in the depolarization level of the specific neurons (i.e., level of excitatory post-synaptic potentials is augmented), increase in the firing frequency of the specific neurons and decrease in the latency with which these neurons begin to discharge. This permits sending their impulses to further levels of information processing in brain in a facilitated mode.

The processing of a newly appearing stimulus is thus serviced by two processes: (1) fast stimulus-specific responses by drivers and (2) a slower, spatially dispersed modulation via the collaterals through the non-specific thalamus. Since the latency of the cortical response to non-specific modulation is considerably slower compared to the afferent latency of the specific cortical neurons (measured in response to the actual stimulus input), the driver-neurons, initially activated only by the specific afference, have to wait for the arrival of the stimulus-related modulatory input. This secondary input has been shown to be necessary for explicit perception (awareness) of the stimulus information pertaining initially to preconscious specific representations (Baars, 1997; Bachmann, 1984, 1994; LaBerge, 1997; Llinás & Ribary,