The binding problem lives on: comment on Di Lollo

Jeremy M. Wolfe

Brigham & Women's Hospital, Harvard Medical School, 64 Sidney St. Suite 170, Cambridge, MA 02139, USA

In his opinion article [1], Vincent Di Lollo argues that the binding problem [2–4] is 'ill-posed'. I disagree. The binding problem is, in fact, a very real problem, solved in the visual system with the aid of selective attention. What was 'ill-posed', or just incorrect, was the traditional mapping of the problem onto the brain.

The binding problem could be described as the problem of knowing how basic features relate to each other. Thus, in Figure 1, it is perfectly clear, once you attend to the upper left-hand item, that it consists of vertical purple and green horizontal regions. Without directing your attention, you can see that the rest of the image contains similar items and you can easily discern that some of those items are rotated 45 degrees. However, you will have no idea if items possess green vertical or horizontal regions unless you attend to them. That is the binding problem. It has not gone away. (Of course, this is a demonstration figure. It may suffer from 'crowding' issues [5], as well as binding problems. However, in the laboratory, the same point can be made with very small numbers of widely spaced items [6]).

Di Lollo challenges an early neurophysiological account of binding, which held that features such as color and orientation were processed in geographically separate pieces of cortex and, thus separated, the features needed to be 'bound' together in order to see coherent objects. He argues that this old 'feature module' idea was over-stated: cells tend to be tuned for multiple features and, therefore, binding is not necessary.

This throws the binding baby out with the physiological bathwater – illustrating, as it does so, the perils of considering neuroscience to be the arbiter of the value of psychological/cognitive models. In fact, it does not make the slightest difference to the reality of the binding problem if color and orientation are handled in separate 'modules' or in the same piece of brain. The fact remains that, whatever your cells may be doing, you, as the user of those cells, cannot distinguish green vertical plusses from green horizontal ones until you selectively attend to one item.

The 'problem' in the binding problem is not geography; it is capacity. The problem is that the nervous system is limited in the number of objects it can recognize at the same time. The limit may be one object or very few but, as you look at the world in front of you, the collections of features do not form themselves into recognized objects until some collection of features, forming something like a 'proto-object' [7], is selected by attention. At that point, the features can be said to be bound into a recognizable item.

Corresponding author: Wolfe, J.M. (wolfe@search.bwh.harvard.edu).

You seem to be able to perceive some aspects of the figure without selective attention. You know, for instance, that these are green and purple plusses. Similarly, it is possible, in the absence of binding, to make some successful decisions in laboratory tasks. These can be quite complex, such as identifying the presence of an animal [8], and may reflect the power of feed-forward processing [9]. Beyond that, however, Di Lollo is probably correct to invoke 'iterative reentrant processing'. If this processing consists of cycles of feedback from higher cortical processes reaching back to make contact with visual information in earlier areas, that sounds a great deal like what might be proposed in 2012 as the physiological substrate of selective attention. Like earlier notions of geographically distinct maps for every feature, this mapping of physiology onto behavior might be incorrect. Eventually, we may learn how the brain implements behavior. While we wait, visual processing will remain capacity limited. In the absence of attention, we will know rather little about how basic features are tied to their objects, and the binding problem will remain well-posed.

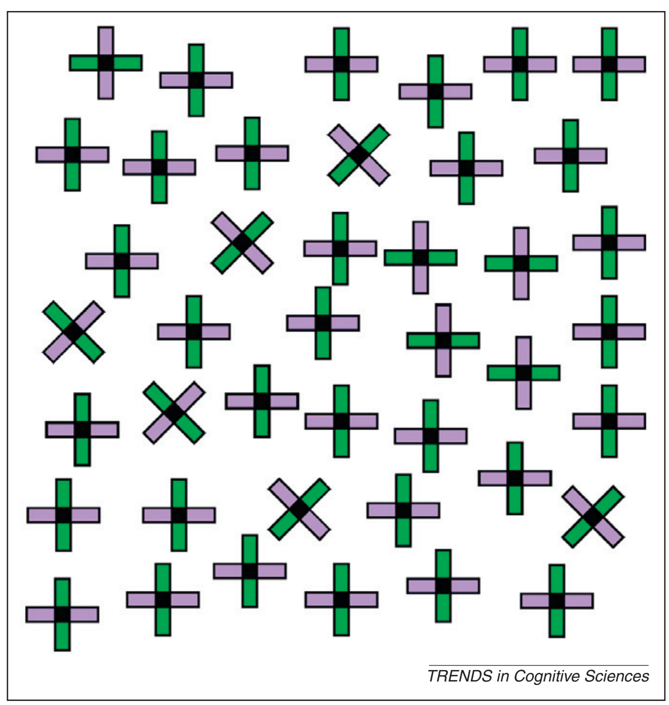

Figure 1. In this figure, it is immediately obvious that there are green and purple 'plusses', some of the rotated to form 'X's. However, it will require attention and binding to determine if a given plus has green vertical or green horizontal components.

References

1 Di Lollo, V. (2012) The feature-binding problem is an ill-posed problem. *Trends Cogn. Sci.* 16, 317–321
2 Malsburg and von der, C. (1981) The correlation theory of brain function. In *Models of Neural Networks II: Temporal Aspects of Coding and Information Processing in Biological Systems* (Domany, J.L. *et al.*, eds), pp. 95–119, (reprinted in1994), Springer-Verlag
3 Treisman, A. (1996) The binding problem. *Curr. Opin. Neurobiol.* 6, 171–178
4 Wolfe, J.M. and Cave, K.R. (1999) The psychophysical evidence for a binding problem in human vision. *Neuron* 24, 11–17
5 Rosenholtz, R. *et al.* (2012) A summary statistic representation in peripheral vision explains. *J. Vis.* 12, article 14
6 Wolfe, J.M. and Bennett, S.C. (1997) Preattentive object files: shapeless bundles of basic features. *Vis. Res.* 37, 25–43
7 Rensink, R.A. (2000) Seeing, sensing, and scrutinizing. *Vis. Res.* 40, 1469–1487
8 Li, F.F. *et al.* (2002) Rapid natural scene categorization in the near absence of attention. *Proc. Natl. Acad. Sci. U.S.A.* 99, 9596–9601
9 Serre, T. *et al.* (2007) A feedforward architecture accounts for rapid categorization. *Proc. Natl. Acad. Sci. U.S.A.* 104, 6424–6429

http://dx.doi.org/10.1016/j.tics.2012.04.013 Trends in Cognitive Sciences, June 2012, Vol. 16, No. 6

Response to Wolfe: feature-binding and object perception

Vincent Di Lollo

Department of Psychology, Simon Fraser University, Burnaby, BC V5A1S6, Canada

The main message of my article [1] was that the feature-binding problem as originally formulated by von der Malsburg [2] was ill-posed. I also asserted that the related problem of object perception is a real problem to which I offered a tentative solution based on iterative reentrant processing. In his commentary [3], Jeremy Wolfe agrees that the original feature-binding problem was ill-posed. He also agrees that object perception is a real problem, but his solution is in terms of binding mediated by attention.

In current usage, 'binding' has become a catch-all term for many things. For example, Treisman [4] has identified no fewer than seven 'binding problems', from 'property binding' to 'conditional binding'. More recently, binding has been said to apply to problems in perception, attention, working memory, and long-term memory [5]. Given this context, attempting to solve 'the' binding problem would be futile, because the solution to any one of those problems would probably not apply to the others. At any rate, using a single term to denote many things can be an impediment to communication and understanding. By means of an operational definition, Wolfe uses the phenomenological appearance of the elements in an image (Figure 1 in [3]) to define the binding problem and to assert that it is still well-posed. I agree that object perception is a real problem but not that it is a binding problem.

What I find difficult to accept is the claim that binding, as evidenced in that image, is mediated by attention. Perception of the stimuli in that image could be mediated just as plausibly by the kind of reentrant processes outlined in my article [1], a possibility acknowledged by Wolfe [3]. This would avoid the use of such a nebulous and ill-defined concept as 'attention'. Just like 'binding', 'attention' has become a catch-all term instantiated in metaphors such as 'spotlight', 'zoom-lens', 'glue', 'limited resource' and 'filter', none of which specifies what mechanism mediates the purported function. So, when 'attention' is invoked to explain 'binding' in Wolfe's image, one is left wondering just what mechanism is involved.

In summary, Wolfe and I agree that the feature-binding problem, as originally formulated by von der Malsburg [2], was ill-posed and that the perception of objects from aggregates of disconnected features is a problem in need of solution. An important point is that the differences in our proposed solutions go beyond issues of mere terminology (it matters little whether the process of correlating a perceptual hypothesis with the ongoing activity at lower cortical regions is referred to as 'attention'). What matters is the way in which the two theoretical convictions would orient the direction of research. In one case we would search for a process (attention?) that selects features and actively binds them into objects. In the other, we would search for a process of correlation, in which those low-level features act as an active blackboard for the perceptual hypotheses sent back from higher regions.

As for Wolfe's misgivings about the 'perils of considering neuroscience to be the arbiter of psychological/cognitive models', I share them, but only up to a point. True, as was the case in von der Malsburg's binding problem, incorrect neurophysiology can lead to incorrect cognitive models. But this does not mean that we can safely ignore neuroscience. For example, although it is entirely sensible for a machine-learning model to assume that a million operations can be performed in one millisecond, we could not use that model to simulate learning in a biological system that is slower by several orders of magnitude. What von der Malsburg's binding mistake teaches us is to be watchful and cautious. Nevertheless, it is still sensible for our models to be guided and constrained by what we know of the system's neuroanatomy and neurophysiology.

Corresponding author: Di Lollo, V. (enzo@sfu.ca).