

## The Binding Problem

Objects have different features such as color, shape, sound, and smell. Some, such as color and sound, are represented separately from the instant they hit our sensory receptors. Other features, such as color and shape, are initially encoded together but subsequently analyzed by separate areas of the brain. Despite this separation, in perception the brain must represent which features belong to the same object. This is the binding problem. Any case of the brain representing as associated two features or stimuli that are initially represented separately can be called binding, but this entry will focus on a subset of these: the pairing of features that belong to a common object.

### Solutions to the spatial binding problem

A simple solution to the binding problem is to have a single neuron (or other representational unit) for each possible combination of features. However, considering that different feature dimensions such as color, shape, and texture may each have hundreds of values, it is impractical to dedicate a unit to each combination. Still, the visual system does contain neurons selective for certain combinations of features, and these may suffice to solve the binding problem in certain cases (Risenhuber & Poggio 1999).

Wolf Singer has championed the theory that binding is represented via synchronous rhythmic firing of the neurons selective for the paired features (von der Malsburg 1981; Gray et al. 1989). The idea is that the joint activity of the feature representations allows other brain areas to process the features together, to the exclusion of features belonging to other objects. Groups of neurons in many parts of the brain frequently do synchronize their responses, and attention to visual stimuli can enhance the effect, but the precise relationship of the phenomenon to perceptual binding remains unclear (Fries et al. 2001; Thiele & Stoner 2003; Dong et al. 2008).

For experimental psychology researchers, two papers on binding by Anne Treisman in the 1980s set the course for nearly two decades (Treisman & Gelade 1980; Treisman & Schmidt 1982). Treisman's "feature integration theory" (FIT) became not only the most influential theory of binding, but also the most influential theory of attention. FIT posits that binding is accomplished by an act of selective attention and has three elements.

- 1) Features like color and shape are represented separately in the brain, but for each feature (such as red) there is a feature map indicating the location of each instance of the feature in the visual field. By virtue of the position of the units that represent the feature relative to the others in the map, a **location tag** is implicitly included and activity of a unit signals both feature identity and location.
- 2) The objects of a visual scene initially cause representations of the various features of all the objects to become active. The system does not yet represent which features belong to the same object.
- 3) Binding happens when **attention** is directed to a particular location. The neurons corresponding to this location in each feature map become active to the exclusion of those in other locations, and the features occupying the location are bound.

### **Spatial Binding: Evidence**

The results of thousands of visual search (see also *visual search*) experiments have been interpreted in the framework of FIT. In one case people were tested on the time needed to find a red 'X' target stimulus among a large array of red 'O's and blue 'X's. Because the target is defined by the combination of color and shape, the task requires binding of these features. If there were no binding to determine whether any instances of red were in the same location as the X shape, the target could not be found. Searches that require binding are usually more time-consuming than searches for a target containing a feature different from those of all the other items. In this latter case, a feature map alone is sufficient to solve the task. According to FIT, the reason for the greater difficulty in the binding search is that the binding step takes time and can only be performed in one location at a time. Attention is hypothesized to visit each location in turn, binding its features until the target is found.

FIT has been useful for explaining visual search results, however a number of other models, models **without any role for attention** in binding, can also explain visual search performance (Rolls & Deco 2002; Eckstein 1998). These alternative models imply that the nature of the binding process cannot be determined from visual search results alone. Visual search results are **affected by many factors such as image segmentation mechanisms, local salience processing, and crowding, making it difficult to isolate the binding process.**

In experiments relying on fewer assumptions, Treisman documented clear binding mistakes with simple displays, in a phenomenon she called "illusory conjunctions" (Treisman & Schmidt 1982). In one illusory conjunctions experiment, two black digits and three colored letters were flashed briefly on the screen. Participants attempted to report the digits and the letters and the letters' associated colors. They usually reported the correct letters and colors, but occasionally with the wrong pairing. Treisman suggested that the short presentation duration, in combination with the attentional demands required to also report the digits, meant that the attention available was sometimes inadequate to correctly bind the features. Subsequent work verified that illusory conjunctions are perceptual errors rather than an artifact of guessing strategies (Ashby, Prinzmetal & Maddox 1996), and illusory conjunctions clearly are more common when perceptual demands are high, but it remains unclear whether this reflects a critical role for attention or instead just that binding requires additional sensory processing beyond that required for identifying the features.

To probe the role of attention more directly, Jochen Braun & colleagues devised a quantitative measure of the attentional resources shared by two tasks (Braun, Koch, Lee & Itti 2001). Results from this paradigm indicate that contrary to FIT, some feature bindings are perceived at little or no cost to attentionally demanding visual tasks. For example, accuracy in reporting the orientation and color of two line segments in the periphery was almost completely unaffected by the demands of a concurrent task of searching for a T among L's or an L among T's. Results from Braun's paradigm do however support the idea that attention is critical for linking features to particular spatial locations. Ability to judge the spatial configuration of adjacent red and green patches (whether red is left of green or right of green) traded off linearly with performance in concurrent central tasks, supporting a critical role for attention. Interestingly, other discriminations that may also rely on configuration do not show this property, for instance discrimination of face gender or identity (Reddy et al. 2004). Apparently the binding required for certain discriminations can proceed with little to no attention. However, even when full attention and extensive processing time is available, binding can fail profoundly, as described in the next section.

### **Location Tagging of Features: A Prerequisite for Binding?**

In her Feature Integration Theory, Treisman suggested that bindings between features are mediated by the **features' links to a location in common**. Psychophysical demonstrations of binding failures under conditions of full attention provide support for this idea that binding is accomplished through common location tags.

In a display devised by Steven Shevell et al. (2008), two separate objects are presented at the same location but in different eyes. For example a vertical array of orange and gray stripes is presented in one eye, and a horizontal array of gray and blue stripes in the other eye. The conflict between the eyes causes experience to alternate between various percepts including some involving misbinding of the color and form features. Specifically, arrays of orange with blue stripes of either orientation are perceived. This phenomenon suggests that **pairing of features is not fully resolved until after the representations of the two eyes come together**. That presentation of multiple instances of a feature in a single location can confound binding supports the notion that binding relies on location tags.

**Even when features are perceived in their correct locations, binding can still fail profoundly.** Still, this too may reflect a location tagging failure that arises when the spatial scale of analysis of one feature is larger than that of another, allowing the location tag of the larger feature to correspond to more than one instance of the smaller feature. In a display devised by Hugh Wilson and Frances Wilkinson (1998), a pair of dots defines a local orientation. Hundreds of such pairs of dots are scattered across the screen and oriented such that globally, a spiral is perceived. If all these dots are white and unrelated black dots are randomly interspersed with them, then under brief presentation conditions binding fails utterly - people are unable to say whether the dots forming the clearly-visible shape are black or white. This phenomenon suggests that the mechanisms for a feature (here, shape) extracted by combining a number of local components does not preserve information about other aspects of its constituents (here, color). Rather, the global shape is assigned to a large area, with multiple colors tagged with locations within it.

Rare neuropsychological syndromes further highlight the role of location tagging. Some patients with bilateral parietal damage mispair color and shape much more frequently than they misperceive the constituent features (Friedman-Hill et al. 1995). Damage to the pulvinar, a

subcortical structure, can cause similar binding problems (Ward et al. 2002). Both disorders are accompanied by a deficit for localization of even an isolated feature, consistent with the thesis that binding of visual features is accomplished through common location tags.

### **Temporal Binding**

Certain binding failures documented in healthy observers suggest that the binding process is quite slow relative to feature identification. Two dot patterns, each forming a different global shape (like those mentioned earlier), each a single color (e.g. red or green) were set in alternation (Clifford, Holcombe & Pearson 2002). The two shapes were constructed such that their shapes could not be determined when the alternation rate exceeded the temporal resolution of the shape identification mechanism. The two global shapes were easily perceived at 15 Hz, implying rapid grouping of the dot pairs and extraction of the global shape, and the color of each dot was perceptually obvious. However, without extensive scrutiny, observers were unable to determine which shape was formed of green dots and which shape formed of red dots unless the patterns alternated slower than 3 Hz. Binding of color and motion also shows a slow limit of less than 3 Hz, even when both features are local (Moradi & Shimojo 2004; Arnold 2005). The alternation of colors and motions causes the display to contain more than one feature of each type in a single perceptual location. With binding based on location tags, this yields binding ambiguity when the displays are alternated fast enough to exceed the temporal resolution of the binding process. The 3 Hz result indicates that binding is slow and requires much more time than does identification of the constituent features.

Even at slow rates, binding of the features that occur together in time may pose problems beyond those faced by spatial binding. For spatial binding, linking features based on common location works well thanks to the large number of spatially organized areas in the visual system. In contrast, no chronotopically-organized visual areas have been found, raising the issue of how features might be tagged temporally. Temporal tagging would be unnecessary if all features were processed in the same amount of time. But features have different sensory latencies and processing times (Schmolesky et al. 1998). Perhaps the perceptual systems have a scheme for tagging the time that features actually occurred in the world, as opposed to when they are identified by the brain (Nishida & Johnston, in press), but this is not yet understood.

### From Perception to Memory and Cognition

Once features are bound by perception, these bindings must be maintained by visual short-term memory if we are to do simple tasks like detect changes in scenes. But visual short-term memory has a very limited capacity, and for moving objects, the pairing of object to features may easily be lost. Jun Saiki (2003) has used a display with a triangle, circle, and square of different colors moving about the screen and briefly disappearing behind occluders. Sometimes the objects switch colors or shapes while they are behind the occluders, but people are very poor at detecting this, even though they do well at detecting a new color or a new shape. Understanding the role of bound object representations versus individual feature representations in real-world behavior requires more investigation of the demands binding places not only on initial perception, but also on sustained attention and memory.

Alex O. Holcombe

November, 2008

University of Sydney

See Also: Grouping, Attention: Selective, Attention: Divided, Attention: Physiological

### References

- Arnold, D. H. (2005). Perceptual pairing of colour and motion. *Vision Res*, 45(24), 3015-3026.
- Ashby, F. G., Prinzmetal, W., Ivry, R., & Maddox, W. T. (1996). A formal theory of illusory conjunctions. *Psychological Review*, 103, 165-192.
- Braun, J., Koch, C., Lee, D. K., Itti, L. (2001). Perceptual consequences of multilevel selection. *Visual attention and cortical circuits*. (pp. 215-241). Cambridge, MA: MIT Press.
- Clifford CWG, Holcombe AO, Pearson J. (2004). Rapid global form binding with loss of associated colors. *Journal of Vision*. 4:1090-1101.
- Dong, Y., Mihalas, S, Qiu, F, von der Heydt, R, & Niebur, E. (2008), "Synchrony and the binding problem in macaque visual cortex", *Journal of Vision* 8: 1-16, doi:10.1167/8.7.30
- Friedman-Hill, S. R., Robertson, L. C., & Treisman, A. (1995). Parietal contributions to visual feature binding: evidence from a patient with bilateral lesions. *Science*, 269(11 August), 853-856.
- Fries, P., Neuenschwander, S., Engel, A. K., Goebel, R., and Singer, W. (2001). Rapid feature selective neuronal synchronization through correlated latency shifting. *Nature Neuroscience* 4(2): 194-200.
- Gray, C. M., König, P., Engel, A. K., and Singer, W. (1989) Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature* 338: 334-337.

Holcombe, AO (2009). The Binding Problem. . In E. Bruce. Goldstein (Ed.), *The Sage Encyclopedia of Perception*. Sage. PRE-PRINT, NOT THE PUBLISHED VERSION

- Hong SW, Shevell SK. (2006). Resolution of binocular rivalry: Perceptual misbinding of color. *Vis Neurosci*, 23:561-566.
- Moradi, F. & Shimojo, S. (2004). Perceptual-binding and persistent surface segregation. *Vision Res*, 44(25), 2885-2899.
- Nishida, S. & Johnston, A. (in press). Time marker theory of cross-channel temporal judgements. In R. Nijhawan & B. Khurana (Eds.), *Problems of Space and Time in Perception and Action*. Cambridge University Press.
- Reddy, L., Wilken, P., & Koch, C. (2004). Face-gender discrimination is possible in the near-absence of attention. *J Vis*, 4(2), 106-117.
- Riesenhuber, M. & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 2(11), 1019-1025.
- Rolls, ET and Deco, G (2002). Computational neuroscience of vision. Oxford: Oxford University Press.
- Saiki, J. (2003). Feature binding in object-file representations of multiple moving items. *J Vis*, 3(1), 6-21.
- Shevell, S. K., St Clair, R., & Hong, S. W. (2008). Misbinding of color to form in afterimages. *Vis Neurosci*, 1-6.
- Thiele, A. & Stoner, G. (2003). Neuronal synchrony does not correlate with motion coherence in cortical area MT. *Nature*, 421(6921), 366-370.
- Treisman, A. & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology*, 14, 107-141.
- Treisman, A. & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology*, 12, 97-136.
- von der Malsburg, C. (1981). The correlation theory of brain function. [Intern. Rep., 81-2, MPI Biophysikalische Chemie, Gottingen, Germany]
- Ward, R., Danziger, S., Owen, V., & Rafal, R. D. (2002). Deficits in spatial coding and feature binding following damage to spatioptic maps in the human pulvinar. *Nature Neuroscience*, 5, 99-100.
- Wilson, H. R. & Wilkinson, F. (1998). Detection of global structure in Glass patterns: implications for form vision. *Vision Res*, 38(19), 2933-2947.