

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

The Hemisphere-specific Processing Linking Visual Perception to Cognition

Permalink

<https://escholarship.org/uc/item/4pr0w4wc>

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 45(45)

Author

Holcombe, Alex O.

Publication Date

2023

Peer reviewed

The Hemisphere-specific Processing Linking Visual Perception to Cognition

Alex O. Holcombe (alex.holcombe@sydney.edu.au)

School of Psychology, University of Sydney
Sydney, NSW 2006 Australia

Abstract

How does cognition engage with the visual world? I make the case that multiple object tracking tasks isolate an object selection process that also applies to unmoving objects. Among other characteristics, the hemisphere specificity of object selection sets it apart from cognitive processing. Tracking is blind in that cognition generally does not know which tracked object is which. Contrary to a recent suggestion, this means that trackers do not function as the labeled pointers thought to be necessary to comprehend language or compute certain spatial relations. Instead, tracking has more in common with stimulus-driven attention, saliency, and featural attention.

Keywords: attention; cognition; perception; tracking; binding

The Visual Architecture of the Brain

Every moment, our retinas respond to a massive amount of information. These retinal signals propagate to brain areas that are largely dedicated to sensory and perceptual processing, and provide massively parallel processing. But our minds are severely limited in their capacity to access certain kinds of visual information. Evidently there is a bottleneck, and to determine how we navigate the visual world, we must understand that bottleneck.

Pylyshyn & Storm (1988) demonstrated that humans can track multiple objects simultaneously in the presence of identical distractors. Findings from thirty-five years of multiple object tracking research provide important insights into how the mind connects with the visual world.

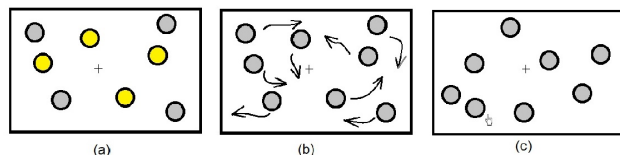


Figure 1: In a typical MOT task, a) targets are highlighted, b) then move about randomly along with the distractors, c) participants must click on the targets' final locations ([Teeeee](#), [CC-BY-SA 4.0](#)).

Hemispheric Doubling

Neurons in the visual stages closest to the eyes operate almost entirely independently of those in the opposite visual hemifield. Low-level visual functions, in other words, are hemifield-independent; indeed, they are local to small individual areas of the visual field. Cognitive processing can be quite different. At the personal level, consciousness is unified – we think of ourselves as unitary persons, rather than

having two independent and simultaneous streams of explicit thought. We are unable to do two long division arithmetic problems at once, and for some simpler laboratory tasks as well, we seem to be so limited in capacity that we can only operate on one object at a time, via a unitary resource rather than duplicate independent processes in the two hemifields (e.g. Oberauer, 2002).

Bridging the enormous processing capacity difference between low-level vision and visual cognition, one might expect a gradual blending rather than a strong boundary. Neuroimaging and neurophysiology have shown that attentional selection involves early retinotopic areas as well as high-level cognitive control areas such as the frontal lobe and parietal cortex. The picture this suggests is one of large parts of the brain cooperating, with heavy involvement of top-down connections. This can be worrying for the prospect of “carving nature at its joints” (Fodor, 1983) to understand the workings of the system. Recurrence is difficult to analyze (Rosen, 2005).

Given this perspective, findings of strong hemifield independence for MOT were a surprise. Alvarez & Cavanagh (2005) found that for a task with multiple moving targets in the left visual hemifield, one can add additional targets to the right visual hemifield at little to no cost. This suggests a strong distinction between the apparently unitary, cross-hemifield functioning of high-level cognition and the processes that link cognition to the visual world.

How independent are the tracking processes (the “trackers”) in the two hemispheres? Three studies yield estimates of independence of 90% or greater (Alvarez & Cavanagh, 2005; Holcombe & Chen, 2012; Chen, Howe, & Holcombe, 2013). Lower estimates may stem from ceiling effects and the use of displays that allow cross-hemifield spatial interference (Holcombe 2023). Another likely factor is a contribution to task performance from the ability of cognitive processes to track a single object.

Overall, it appears that there truly is an almost hemifield-specific process, albeit one that is not trivial to isolate. This suggests that tracking can be thought of as a “module” that receives instructions from cognition but does not demand much cognitive resource.

It's Spatial Selection in General

As MOT is just one particular task, skepticism is understandable of the proposition that its characteristics belong to a critical link between perception and cognition.

Pylyshyn did work showing that tracking multiple objects goes together with other measures of having attention on those objects (Sears & Pylyshyn, 2000). More recent work

has capitalized on the hemisphere specificity of MOT to investigate when the associated processes are critical to attention. Alvarez, Gill, and Cavanagh (2012) found evidence that spatial selection is hemisphere-specific, as expected if it shares processes with tracking.

Spatial selection here means indexing multiple objects based on their locations, in the presence of distractors not easily distinguished by color or another feature. Distractors must be present because without them, selection can occur via stimulus-driven attention, which operates without a limited-capacity resource. Stimulus-driven salience already begins to emerge in early visual cortices, through local differencing and suppression processes.

The reason that the objects should not be distinguishable by a simple feature, is that in such circumstances feature attention can be used to select targets, and feature attention is a distinct and global resource. Participants are unable, for example, to confine their attempts to attend to red items to a particular spatial area (White & Carrasco, 2011; Lo & Holcombe, 2014; Saenz, Buracas, & Boynton, 2002). Featural attention may thus operate via a high-level cognitive instruction, operating via dorsal frontoparietal networks to control feature-based selection by modulating sensory representations in visual areas in a manner is unavoidably sent to visual cortices in both hemispheres to boost salience (Liu, 2019).

To study spatial selection, Alvarez, Gill, & Cavanagh (2012) conducted a series of visual search studies. They compared performance when the stimuli were arrayed bilaterally to when the stimuli were confined to one hemifield. When a rather standard search task was used, they did find an advantage of the bilateral display, but it was small. They then tested a subset search task wherein there were many distractors present, but the participants were informed that the target would occur in one of several locations that they were informed of at the beginning of each trial. In these circumstances a large advantage was observed when the relevant locations were divided between the hemifields. Finally, when the relevant stimuli were specified by color (feature attention) rather than by locations, the hemifield advantage did not occur.

These results support the suggestion that spatial selection is just like MOT in being hemisphere-specific. Ideally we'd be able to quantitatively compare the hemifield-specificity found by Alvarez et al. to that found in MOT tasks. However, the response time measure used in the visual search tasks makes for an apples-and-oranges situation for comparing to the accuracy or speed thresholds used for MOT tasks.

Visual Working Memory Isn't Split

Delvenne and colleagues documented an advantage for dividing stimuli to be remembered among the two hemifields (e.g. Delvenne, 2012; Holt & Delvenne, 2015). However, the advantage is quite small relative to that for MOT (and some don't see an effect, Bays, personal communication) and may be specific to location and shape (p.78 of Umemoto et al., 2010). This suggests that the observed hemifield advantage

may be simply inherited from the location selection process, but is diluted by unitary processes responsible for most of the memory task.

A Flexible Resource

Pylyshyn (1994) suggested that MOT was mediated by "about four" discrete pointers. In the intervening decades, many researchers have continued to claim that the average human brain has a specific limit on the number of objects that can simultaneously be tracked (four and five are most frequently invoked), but the evidence does not support this (Ch. 3 of Holcombe, 2023).

The discrete entities or pointers conception of what provides task capacity is referred to as "slots" in the visual working memory literature. The alternative of an undifferentiated resource pool that can be divided among as many objects as one likes is referred to as "resource theory". Visual working memory researchers have devised many experiments and sophisticated mathematical analyses to distinguish between slots and resources. The debate over which best characterizes visual working memory is not over, but it seems difficult for a discrete slot theory to explain all results (Bays, Schneegans, Ma, & Brady, 2023).

In MOT as well, there is no clear evidence for slots, despite it being the favored theory for decades (Ch. 3 of Holcombe, 2023). The efforts to investigate the issue have not settled the debate, but they have made clear that people can apply the tracking entities (whether they come from slots or resources) very flexibly. It is not clear whether this happens via simultaneously sharing a resource among several targets or time-sharing one or a few trackers among them.

The degree of flexibility and robustness of tracking seems very different from explicit central cognitive processes. For example, if you were given two long-division problems, you'd be hard pressed to do both of them at the same time in the same way that people can track four targets with the same ease as two targets, albeit with lower performance. Tracking is also surprisingly robust to dual-task demands (e.g. Thornton & Horowitz, 2015), further setting it apart from central cognition.

Selection is Blind

What Pylyshyn wanted from spatial or object selection is the functionality provided by the addressable pointers used in digital computers. Indeed, Pylyshyn was inspired by this aspect of digital computers to devise his experiments demonstrating that people can track multiple objects.

For flexible computation, computers require pointers to addresses in memory. These are conspicuous in programming languages such as C but are abstracted away in higher-level languages such as Python.

Two aspects of pointers are crucial. One is that they can point to lots of things, so the mere existence of a pointer does not indicate anything about what is out there. A second aspect is that pointers are kept distinct and *indexed* (Pylyshyn, 2007). By indexed, I mean that you know which one is which, so they can be assigned different roles.

A programmer who knows they will need to calculate a function of two numbers, x , and y , on multiple occasions (with different values for x and y on the different occasions) might dedicate pointer A to pointing at x and pointer B to pointing at y . Every time $f(x,y)$ needs to be calculated, pointer A is set to point at wherever the new x is in memory, and pointer B is set to point at wherever the new y is in memory. The reason that knowing which is which (indexing) is important is that $f(x,y)$ may not equal $f(y,x)$; if you don't know which pointer is A and which B, you could get the wrong answer. A serves a different role than B.

The ability to distinctly refer to items that different roles is also important in language production and comprehension (Frege 1980). Language has syntactical structure, such that placing words in different locations in a sentence changes the meaning of the sentence (“Swim goggles” means something quite different than “Goggles swim.”). Comprehending a sentence, then, involves knowing the different roles particular positions in a sentence are associated with. “Compositionality” is a popular term for this (Frankland & Greene, 2020).

To represent structure, a pointer-like system, with its property of role-filler independence (Hummel et al., 2004), is handy. One pointer can be assigned to one structural role (e.g., the actor) and another pointer to a different structural role (e.g., the recipient). The content these point to would be the corresponding words for a particular utterance. A brain or computer can then have a rule to begin the process of comprehending a sentence based on processing the words in different roles differently. The rule is applied to the sentence via the pointers.

While these concepts may fit more naturally to a von Neumann computer architecture than to that of the brain, neural network models have also been devised that mediate the structure-content and pointers division of mechanisms (Russin et al., 2020). It has also been suggested that the dorsal stream specializes in structure and pointers with the ventral stream specializing in content (Frankland & Greene, 2020).

Pylyshyn laid out his FINST theory of tracking largely to explain cognitive abilities that require indexed pointers. Recently, O'Reilly, Ranganath, & Russin (2022) picked up on this for their theory of how neural networks give rise to flexible intelligence, citing the FINST theory of indexed tracking approvingly, e.g. “The FINST (fingers of instantiation) framework of Pylyshyn (1989) provides an early, simple model for how structure can be represented independently of specific content and also maps well onto cognitive-neuroscience data reviewed in the next section.” and “FINST-like indexes provide a plausible attention-based neural mechanism for role-filler variable slots in the context of classical symbolic representational frameworks.” (p.125)

A major problem, however, with this ambit for selection is that the processes attached to each tracked object do *not* seem to be indexed. The pattern of task performance suggests that when people track multiple objects, they are typically oblivious to what it is they are tracking. While the processes that mediate tracking frequently update their representations

of the locations of targets, they seem to rarely inform cognition of what those objects are or what features they have.

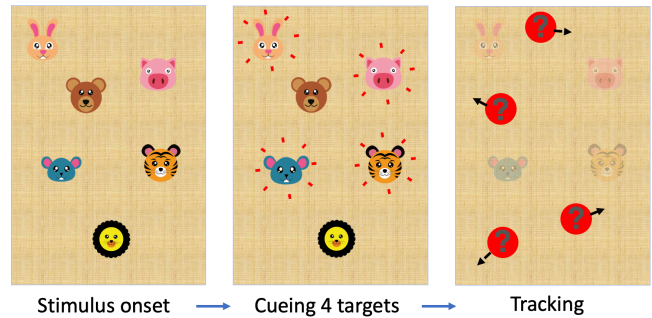


Figure 2: Asked to keep track of the identify of moving targets, participants quickly lose track of which is which. It is as if the identities are left at the starting location where the targets were first memorized (Hollingworth & Rasmussen, 2010).

This supports the arbitrariness aspect of pointers (they continue pointing at a thing regardless of what it becomes) but appears to prevent their potential use as syntactic pointers. Holcombe (2023, Ch. 10) reviews the evidence for this blindness of pointers. While it has been suggested that tracking processes may need to be distinct to link successive locations of a target to each other rather than to another target (Pylyshyn, 2004), any representation of which object is which does not extend to explicit knowledge of which is which.

One effect of the lack of featural updating is that when bottom-up saliency does not draw selection to an individual object, people are blind to dramatic changes. Saiki & Holcombe (2012) demonstrated this with a large field of dots. When the dots were stationary, any change in the color of a dot was easily detected. But when the dots were all in motion, people were blind to such changes. Suchow & Alvarez (2011) created a dramatic demonstration based on this principle.

When Guido van Rossum created the Python programming language, he included a data structure called a set. Sets are unordered lists, so one can check whether a set contains a particular item, and one can add items to a set, but a set does not provide indexed pointer functionality – much like our MOT processing.

I hasten to add that the evidence is not currently strong enough to claim that we are *totally* blind to what we track. One problem for garnering such evidence is the existence of focused attention, which is more cognitively mediated and allows us to frequently update our knowledge of the features of a single target. Any assessment of the amount of blindness of the MOT mechanisms needs to take the potential “contamination” by focused attention into account.

The blindness of tracking prompts the question of whether people even know how *many* objects they are tracking. I don't know of any data on this. A good test would need to use more than four or five targets, otherwise people could enumerate (subitize) the targets so rapidly that one could scarcely detect

that they didn't know their number beforehand. If it turns out they don't know, then tracking is even more blind – like a black box that you can put visual objects in but one for which you can easily forget how many items it contains.

Filling the Slots Elsewhere

Pylyshyn was surprised by the evidence that tracking uses blind, unindexed pointers. That finding didn't satisfy the desire for distinct placeholders that one can structure a group of items, facilitating the application of rules, such as for language. It's unfortunate for spatial cognition, too. Shimon Ullman (1984) pointed out that we often need to make spatial judgments that require the sort of flexibility that precludes using a template. For example, judging whether one object, A, is inside a larger shape, B, or perhaps whether an object C is above, below, to the left, or the right of an object D. Because the answer to whether A is inside B can be different than the answer to whether B is inside A, we need to keep track of which object is A and which B.

The non-indexed nature of trackers and their hemisphere specificity marks them as quite distinct from cognition. This is good for science in that humans are most successful at understanding things that can be broken into parts (Simon, 1969), but it does mean that cognition does not have trackers as a module to represent structure as Pylyshyn had hoped.

The blindness of tracking also indirectly supports the claim that perception is informationally encapsulated from cognition, as championed by Fodor (1983) and more recently reviewed by Firestone & Scholl (2016).

If the processes underlying tracking don't provide indexing, what does? In a review of the literature, Frankland & Greene (2020) focus on the “default mode network” collection of dynamically-connected brain regions. They consider a few representational schemes varying in how distributed they are and how much separation there is between roles and fillers, including grid cells and “map-like representations” interacting with superior temporal cortex and the prefrontal cortex.

While Frankland & Greene (2020) highlight a number of brain areas as potentially important, based on a range of language comprehension and other tasks, they notably do not highlight the parietal regions most consistently activated by MOT (in particular as having activation that increases with number of targets, that helps isolate the tracking component of the task, e.g. Culham, Cavanagh, & Kanwisher 2001). The most related region they highlight appears to be the angular gyrus, whereas MOT regions are more superior and anterior in parietal cortex.

What role are multiple object tracking processes left with? Although they are not as involved in the amodal, often linguistic tasks reviewed by Frankland & Greene (2020), parietal regions activated by MOT are critical for apprehending visual spatial relations and calculating coordinates for action. Tracking processes may be key to these abilities, and more abstract amodal representations may be created subsequently by other areas.

Using Your Blind Trackers

If tracking operates largely blind, which may mean that location selection broadly is blind, how can we apprehend spatial relations? We have already said that MOT makes spatial layout, the positions occupied by the targets, available. This seems to reflect parallel access by cognition to the trackers' spatial locations. The problem arises when one wants to know, say, whether a particular target, say a red target, is to the left or to the right of a green target. A line of experimentation completely independent of MOT has reached the same conclusion, culminating in “Boolean map theory” (Huang, 2010), according to which attention can represent multiple locations but not multiple values of other features.

How can we apprehend multiple non-location features? For some tasks, good guesses can be made when motion patterns are somewhat predictable, as frequently happens in real-world situations such as the sports field or when tracking children at the playground. But in a laboratory experiment devoid of that sort of predictability, it appears that focused attention must be serially deployed to the two targets (setting aside when attention to an individual feature value is useful to reveal the locations of that particular feature value). This time-consuming process is required to deliver to cognition the information of what color is at a particular location (e.g. Treisman, 1980).

Using two concentric arrays of moving colored discs, Holcombe, Linares, & Pashkam (2011) asked participants to report what color in the outer array was spatially aligned with a particular color in the inner array. The pattern of results indicated that participants first identified the object with the cued color in the inner ring and then shifted their attention to its partner in the outer array. In particular, when the objects moved fast enough that people sometimes made errors, their errors were most frequently the color in the outer array that trailed the correct color, as if participants shifted attention but sometimes the target had moved on and the shift landed on the lagging object. Moreover, when the objects moved too fast to track at all, the colors could still be perceived (perhaps through feature attention) but their spatial relations could not be determined. Using different techniques, Franconeri et al. (2012) arrived at the same conclusion that focused attention is necessary to determine the spatial relationship of two features.

The Contribution of Stimulus-Driven Attention

Multiple object tracking tasks appear to reveal the characteristics of top-down location selection in general, not just that for unmoving static objects. Tracking moving objects appears to reflect the same processing as top-down spatial selection of unmoving objects.

But locations can also be selectively processed as a result of stimulus-driven attention (in MOT, while objects may be flashed initially such that targets are acquired via stimulus-driven attention, after the targets become identical to the distractors, selection has to be maintained by top-down selection).

Just as in top-down location selection, it appears that the objects linked to cognition via featural attention and stimulus-driven attention aren't indexed. For stimulus-driven attention, for example, if several stimuli are equally salient as a result of a sudden onset (stimulus-driven attention), attention is drawn to them, but not in a way that delivers all of their identities, or indexes them. Indeed, saliency is often conceptualized as being represented by a "saliency map" that is unlabelled (Niebur, 2007).

For featural attention as well, while one can be fairly confident that if one sets out to attentionally select red, all the items then selected will be red, this doesn't deliver information about what those objects are nor does it index them (Huang, 2010). Still, one's knowledge that the act of selection was initiated by attention to red likely contributes to the creation of central representations that include non-location feature values (here, color) as well as location.

Bottom-up, stimulus-driven attention likely contributes not only to the initiation of selection when objects are highlighted at the beginning of an MOT trial. It likely also contributes to maintenance of selection on moving objects during MOT (Holcombe, 2023, p. 30). When I spatially select some unmoving objects, and then they begin moving, my phenomenology is consistent with my attention being pulled along by the objects. Tracking feels almost automatic. If I instead attempt to keep my attention fixed to the original locations of the objects as they move, this feels unnatural and difficult.

The spatiotemporal filters of the visual cortex that are driven by the sudden onsets in attentional capture paradigms are approximately equally activated by the movement of an object. With spatial attention already on the targets, on each frame their movement elicits neural activation in a contiguous set of locations, so attention is naturally dragged with the targets. No direct evidence has been provided for this, but it illustrates how tracking can be rather parsimoniously and intimately related to conceptions of attention derived from the literature on unmoving but suddenly-onsetting objects.

Seething Saliency

As laid out in the previous section, selection happens as a result of bottom-up stimulus-driven saliency as well as the top-down factors of location selection and featural attention. While the resources that determine top-down location selection are hemisphere-specific, featural attention may not be and bottom-up saliency is highly local. The end result, however, may be just different degrees of saliency. This is quite different from Pylyshyn's conception of discrete pointers that either are on or off, without different degrees. Selection may be more seething (continuously varying) than symbolic. The existence of retinotopic maps and multiple mechanisms to modulate the activity of neurons may mean that the brain can easily vary its prioritization of many locations in parallel, whereas making a connection to non-location features requires an extra step.

There are various dissociations between the perception of spatiotemporal relations and the perception of non-location features, and most go in the direction of perception of spatial or temporal location without knowledge of the non-location feature. We have already mentioned the blindness to non-location features evident in multiple object tracking displays, of course. In the temporal domain, judgments of whether two things occur at different times can be made at fast stimulus presentation rates, but these fast rates prevent identification of which feature belongs to which stimulus (e.g. Holcombe & Cavanagh, 2001). Finding a similar dissociation in multiple modalities, not just vision, Fujisaki & Nishida (2010) suggested that apprehending the non-location features of stimuli typically requires an extra processing step.

For color and shape information, there may be other routes to cognition than focused attention and featural attention. For example, there is an ongoing debate about simultaneous awareness throughout the visual field and the status of "statistical perception" of ensembles of objects (Whitney & Leib, 2018). This ability may not provide much information about the layout of what is perceived, but this requires further study – I do not know of work that has established the relationship to the serial attention that other evidence points to the need for.

Conclusion

Pylyshyn and other theorists expected a strong mechanistic link between attentional selection and awareness of features. But it seems that the brain doesn't work that way. We can't keep straight which of a set of moving objects is which, let alone maintain their other features in our minds. The additional neural activation caused by saliency or selection does cascade into other brain regions, however, and it is with those that object representations can become more robust and labelled. This may be particularly true of locations that are selected intentionally, in part because it being intentional usually means that cognitive processes already had some representation associated with those locations.

Selection's consequence of visual activation carrying on into later brain regions is important, but this may lead one to imagine a continuous cascaded system of mass action without strong division of function. That would be a misleading image, because assigning features to locations usually has to occur one-by-one via serial attention (apart from an indirect route of feature attention mentioned earlier). In addition, certain aspects of location selection are hemifield-specific, and the reason why remains unknown, although we can speculate it reflects separate pools of neurons in the left and right hemispheres of parietal cortex.

The experimental methodologies that have been developed over the past few decades have revealed a hemisphere-specific bottleneck that is critical plays an important role in selecting locations for further processing. These methodologies should be built on to better isolate the subsequent processing that yields most of our knowledge of what is where.

Acknowledgments

I thank the anonymous reviewers for comments, and Paul Bays and Sebastian Schneegans for discussion.

References

- Alvarez, G. A., & Cavanagh, P. (2005). Independent resources for attentional tracking in the left and right visual hemifields. *Psychological Science*, *16*(8), 637–643.
- Delvenne, J. (2012). Visual short-term memory and the bilateral field advantage. In Kalivas, G and Petralia, SF, editors, *Short-Term Memory: New Research*. Nova.
- Fendrich, R., Wessinger, C. M., & Gazzaniga, M. S. (1996). Nasotemporal overlap at the retinal vertical meridian: Investigations with a callosotomy patient. *Neuropsychologia*, *34*(7), 637–646.
- Firestone, C., & Scholl, B. J. (2016). Cognition does not affect perception: Evaluating the evidence for “top-down” effects. *Behavioral and Brain Sciences*, *39*.
- Fodor, J. A. (1983). *The Modularity of Mind*. MIT Press, Cambridge, MA.
- Frankland, S. M., & Greene, J. D. (2020). Concepts and compositionality: In search of the brain’s language of thought. *Annual Review of Psychology*, *71*, 273–303.
- Francis, G., & Thunell, E. (2022). Excess Success in Articles on Object-Based Attention. *Attention, Perception & Psychophysics*, *84*.
- Franconeri, S. L., Scimeca, J. M., Roth, J. C., Helseth, S. a, & Kahn, L. E. (2012). Flexible visual processing of spatial relationships. *Cognition*, *122*(2), 210–227.
- Frege G. 1980. Letter to Jourdain. In *Philosophical and Mathematical Correspondence of Gottlob Frege*, ed. B McGuinness, pp. 78–80. Chicago: Univ. Chicago Press
- Holcombe, A. (2023). *Attending to Moving Objects*. Cambridge University Press.
- Holcombe, A. O., & Cavanagh, P. (2001). Early binding of feature pairs for visual perception. *Nature Neuroscience*, *4*(2), 127–128. <https://doi.org/10.1038/83945>
- Hollingworth, A., & Rasmussen, I. P. (2010). Binding objects to locations: The relationship between object files and visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(3), 543.
- Holt, J. L. and Delvenne, J.-F. (2015). A bilateral advantage for maintaining objects in visual short term memory. *Acta Psychologica*, *154*:54–61.
- Hummel JE, Holyoak KJ, Green C, Doumas LA, Devnich D, et al. 2004. A solution to the binding problem for compositional connectionism. In *Compositional Connectionism in Cognitive Science: Papers from the AAAI Fall Symposium*, ed. SD Levy, R Gayler, pp. 31–34. Palo Alto, CA: AAAI.
- Huang, L. (2010). What is the unit of visual attention? Object for selection, but Boolean map for access. *Journal of Experimental Psychology: General*, *139*(1), 162–179.
- Liu, T. (2019). Feature-based attention: Effects and control. *Current Opinion in Psychology*, *29*, 187–192.
- Niebur, E. (2007) Saliency map. *Scholarpedia*, *2*(8):2675.
- Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3):411.
- O’Reilly, R. C., Ranganath, C., & Russin, J. L. (2022). The structure of systematicity in the brain. *Current Directions in Psychological Science*, *31*(2), 124–130.
- Pylyshyn, Z. W. (2007). *Things and Places: How the Mind Connects with the World*. MIT Press.
- Pylyshyn, Z. (2004). Some puzzling findings in multiple object tracking: I. Tracking without keeping track of object identities. *Visual Cognition*, *11*(7), 801–822.
- Pylyshyn, Z., Burkell, J., Fisher, B. et al. (1994). Multiple parallel access in visual attention. *Canadian Journal of Experimental Psychology/Revue Canadienne De Psychologie Expérimentale*, *48*(2):260.
- Pylyshyn, Z. W. and Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision*, *3*(3):179–197.
- Rosen, R. (2005). *Life Itself: A Comprehensive Inquiry Into the Nature, Origin, and Fabrication of Life* (1st edition). Columbia University Press.
- Russin, J., Jo, J., O’Reilly, R. C., & Bengio, Y. (2020). Systematicity in a recurrent neural network by factorizing syntax and semantics. In S. Denison, M. Mack, Y. Xu, & B. C. Armstrong (Eds.), *Proceedings for the 42nd Annual Meeting of the Cognitive Science Society* (pp. 109–115). Cognitive Science Society.
- Saiki, J. and Holcombe, A. O. (2012). Blindness to a simultaneous change of all elements in a scene, unless there is a change in summary statistics. *Journal of Vision*, *12*:1–11.
- Scholl, B. (2001). Objects and attention: The state of the art. *Cognition*, *80*(1/2), 1–46.
- Sears, C. R., & Pylyshyn, Z. W. (2000). Multiple object tracking and attentional processing. *Canadian Journal of Experimental Psychology*, *54*(1), 1–14.
- Simon, H. A. (1969/2019). *The Sciences of the Artificial, reissue of the third edition with a new introduction by John Laird* (3rd edition). MIT Press Academic.
- Suchow, J. W. and Alvarez, G. A. (2011). Motion Silences Awareness of Visual Change. *Current Biology*, *21*(2): 140–143.
- Treisman, A., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology*, *12*, 97–136.
- Ullman, S. (1984). Visual routines. *Cognition*, *18*, 97-159.
- Whitney, D., & Yamanashi Leib, A. (2018). Ensemble perception. *Annual Review of Psychology*, *69*, 105–129.