Rosenholtz, R. (2017). Those pernicious items. *Brain & Behavioral Sciences, 40,* e154.
In response to: Hulleman, J. & Olivers, C. N. L. (2017). The impending demise of the item in visual search. *Brain & Behavioral Sciences, 40,* e132.
(preprint)

Those pernicious items

doi:10.1017/S0140525X16000248, e154

Ruth Rosenholtz
Department of Brain & Cognitive Sciences, CSAIL, Massachusetts Institute of Technology, Cambridge, MA 02139.
rruth@mit.edu
http://persci.mit.edu/people/rosenholtz

Abstract: Hulleman & Olivers (H&O) identify a number of problems with item-based thinking and its impact on our understanding of visual search. I detail ways in which item-thought is worse than the authors suggest. I concur with the broad strokes of the theory they set out, and also clarify the relationship between their view and our recent theory of visual search.

Our impression of a scene usually includes objects and their properties. When crossing the street, we consider the location and speed of a nearby car. However, just because we recognize "things" at the output of perception and employ high-level reasoning about objects, this does not mean that our visual systems operate on presegmented things. This is a common and tempting cognitive error, which can hamper uncovering the true mechanisms of vision.

Objects make for a useful abstraction. It is natural, therefore, that many theories of vision describe processes as operating on objects and their features. For example, preattentive vision has been depicted as encoding *item* locations and features. (This is distinct from knowing *image* features, such as the outputs of V1 cells.) According to this view, search is slow because serial selective attention is necessary to bind those features together. Such word models enable easy intuitions and guide new experiments. Furthermore, abstracting from the image input to things and their features can sometimes make modeling tractable, as with signal detection theory.

The authors argue that a focus on items has tainted our ideas about search: it has hampered understanding search in real-world images, for which the set size (number of items) is ill-defined (Rosenholtz et al. 2007; Wolfe et al. 2011a; Zelinsky 2008); it has led to a focus on selective attention as the limiting mechanism, discounting the role of eye movements; it caused the field to focus on the easy end of the performance spectrum; and it has led to over-estimation of the importance of item location. I would argue that thinking about items is even more pernicious.

Item-based theories have not merely biased our choice of stimuli by limiting use of real-world images. Experimenters often design stimuli to preserve the preeminence of the item. One must avoid alignment, which might produce perceptual groups, or else risk violating assumptions that the items can be treated independently. This is analogous to

visual short-term memory experiments that seem designed to give the subject little choice but to remember *items*; it should not surprise us when slot models do well.

Relatedly, only a handful of experiments have studied the effects of image transformations on search: What if we make the items larger or the bars thinner, change the sign of contrast, add noise to the display, or make the displays more dense? Unless these transformations interfere with item visibility, none should have an impact if items are the atoms of search. Yet there is evidence that such transformations do have a significant effect (e.g., Beck et al. 1987; Chang & Rosenholtz 2014; Graham et al. 1992; Rubenstein and Sagi 1996).

More broadly, it is risky to think of the visual input as consisting of an array of items with particular experimenter-defined features. Vertical rectangular bars also contain horizontal edges; oblique filters will also respond to those bars; the "white space" between items also has features; and some features of the display may have a scale larger than any individual items.

The dominance of item-based theories has led to a serious disconnect between theories that essentially operate on experimenter-labeled stimuli (items and their nominal features) and those that operate on actual images. In working with real images, a number of reasonable search strategies do not require items as such, for example applying a template throughout the image and looking for locations with a strong response (Zelinsky 2008). If one does attempt to implement item-based theories, it quickly becomes clear that neither segmenting the items nor determining their supposed "features" is trivial. One is left with the puzzle of why one is "allowed" to use bound features to "preattentively" segment the image into items, but not to recognize the target.

It is hard not to think in terms of items. Despite their main thesis, Hulleman & Olivers (H&O) suggest that target-distractor discriminability is important for setting the size of the functional visual field (FVF). Why discriminability of the *items*? This leaves the puzzle of why search asymmetries abound, as surely target-distractor discriminability is generally the same as distractor-target discriminability. We argue that the major determinant of search performance is crowding (*not* retinal resolution), which demonstrates that peripheral vision operates over sizeable patches, typically containing multiple items. Discriminability of patches is what matters in this scheme (Rosenholtz et al. 2012b).

The authors allude to one theory attributing crowding to limited attentional resolution (Intriligator & Cavanagh 2001). This is subtly item-centric, presuming that attention aims to select only a single item. We have argued this is not ideal in real images (Rosenholtz & Wijntjes 2014). Other theories of crowding also describe mechanisms that operate on items (Greenwood et al. 2009; 2012; Levi & Carney 2009; Parkes et al. 2001; Põder and Wagemans 2007; Strasburger 2005; van den Berg et al. 2012). Recently we have shown that a number of results used to test these item-based theories can instead be explained by the information available in a rich set of image statistics (Keshvari & Rosenholtz 2016). These same statistics plausibly underlie scene perception (Ehinger & Rosenholtz, 2016; Rosenholtz et al. 2012a), suggesting a single encoding scheme could both extract the scene context and support search, in agreement with H&O.

The target article presents clever and thoughtful critiques of prevailing theories, and a new model. The parallels to recent work in my lab are fairly clear (Rosenholtz et al. 2012b; Zhang et al. 2015), though differences raise important questions. We agree that

search likely involves parallel processing, punctuated by serial shifts of the point of fixation. Peripheral vision limits the information available at a glance. Our view is that, rather than being a mechanism, the FVF might *describe* the more informative image regions. It would degrade smoothly, with some regions providing more information than others. It need not be continuous; eccentric uncrowded regions might provide more information than closer crowded ones. The authors are somewhat unclear on these points: Does the FVF have hard edges, outside of which no information exists for telling apart the target and distractor? Is it mechanistic or descriptive? Does some mechanism set the size? If so, how, and why?

References

Beck, J., Sutter, A. & Ivry, R. (1987) Spatial frequency channels and perceptual grouping in texture segregation. *Computer Vision, Graphics, and Image Processing* 37(2):299–325.

Chang, H. & Rosenholtz, R. (2014) New exploration of classic search tasks. *Journal of Vision* 14(10):933.

Ehinger, K. A. & Rosenholtz, R. (2016). A general account of peripheral encoding also predicts scene perception performance. *Journal of Vision* 16(2):13.

Graham, N., Beck, J. & Sutter, A. (1992) Nonlinear processes in spatial-frequency channel models of perceived texture segregation: Effects of sign and amount of contrast. *Vision Research* 32(4):719–43.

Greenwood, J. A., Bex, P. J. & Dakin, S. C. (2009) Positional averaging explains crowding with letter-like stimuli. *Proceedings of the National Academy of Sciences of the United States of America* 106(31):13130–35.

Greenwood, J. A., Bex, P. J. & Dakin, S. C. (2012) Crowding follows the binding of relative position and orientation. *Journal of Vision* 12(3):1–20.

Intriligator, J. & Cavanagh, P. (2001) The spatial resolution of visual attention. *Cognitive Psychology* 43:171–216. doi:10.1006/cogp.2001.0755.

Keshvari, S. & Rosenholtz, R. (2016) Pooling of continuous features provides a unifying account of crowding. *Journal of Vision*, 16(3):39, 1-15.

Levi, D. M. & Carney, T. (2009) Crowding in peripheral vision: Why bigger is better. *Current Biology* 19(23):1988–93.

Parkes, L., Lund, J., Angelucci, A., Solomon, J. A. & Morgan, M. (2001) Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience* 4(7):739–44.

Põder, E. & Wagemans, J. (2007) Crowding with conjunctions of simple features. *Journal of Vision* 7(2):23.1–12.

Rosenholtz, R. & Wijntjes, M. (2014) Peripheral object recognition with informative natural context. *Journal of Vision* 14(10):214.

Rosenholtz, R., Li, Y. & Nakano, L. (2007) Measuring visual clutter. *Journal of Vision* 7(2):17, 1–22. doi:10.1167/7.2.17.

Rosenholtz, R., Huang, J. & Ehinger, K. A. (2012a) Rethinking the role of top-down attention in vision: Effects attributable to a lossy representation in peripheral vision. *Frontiers in Psychology* 3(February):13. doi:10.3389/fpsyg.2012.00013

Rosenholtz, R., Huang, J. Raj, A., Balas, B. J. & Ilie, L. (2012b) A summary statistic representation in peripheral vision explains visual search. *Journal of Vision* 12(4):14. 1–17. doi: 10.1167/12.4.14.

Rubenstein, B. S. & Sagi, D. (1996) Preattentive texture segmentation: The role of line terminations, size, and filter wavelength. *Perception & Psychophysics* 58(4):489–509.

Strasburger, H. (2005) Unfocused spatial attention underlies the crowding effect in indirect form vision. *Journal of Vision* 5(11):1024–37.

van den Berg, R., Johnson, A., Martinez Anton, A., Schepers, A. L. & Cornelissen, F. W. (2012) Comparing crowding in human and ideal observers. *Journal of Vision* 12(6):13. doi:10.1167/12.6.13

Wolfe, J. W., Alvarez, G. A., Rosenholtz, R., Kuzmova, Y. L. & Sherman, A. M. (2011a) Visual search for arbitrary objects in real scenes. *Attention, Perception, & Psychophysics* 73:1650–71.

Zelinsky, G. J. (2008) A theory of eye movements during target acquisition. *Psychological Review* 115:787–835. doi:10.1037/a0013118

Zhang, X., Huang, J., Yigit-Elliott, S. & Rosenholtz, R. (2015) Cube search, revisited. *Journal of Vision* 15(3):9, 1-18. doi:10.1167/15.3.9