Current Biology *21*, 1135–1139, July 12, 2011 ©2011 Elsevier Ltd All rights reserved DOI 10.1016/j.cub.2011.05.031

Report

Perceiving Spatial Relations via Attentional Tracking and Shifting

Alex O. Holcombe,[1],* Daniel Linares,[2]
and Maryam Vaziri-Pashkam[3]
[1]School of Psychology, University of Sydney, Sydney,
NSW 2006, Australia
[2]Perception and Emotion Research Group, Human and
Information Science Laboratory, NTT Communication Science
Laboratories, 3-1, Morinosato-Wakamiya,
Atsugi-shi 243-0198, Japan
[3]Department of Psychology, Harvard University,
Cambridge, MA 02138, USA

Summary

Perceiving which of a scene's objects are adjacent may require selecting them with a limited-capacity attentional process. Previous results support this notion [1–3] but leave open whether the process operates simultaneously on several objects or proceeds one by one. With arrays of colored discs moving together, we first tested the effect of moving the discs faster than the speed limit for following them with attentional selection [4]. At these high speeds, participants could identify which colors were present and determine whether identical arrays were aligned or offset by one disc. They could not, however, apprehend which colors in the arrays were adjacent, indicating that attentional selection is required for this judgment. If selection operates serially to determine which colors are neighbors, then after the color of one disc is identified, attention must shift to the adjacent disc. As a result of the motion, attention might occasionally miss its target and land on the trailing disc. We cued attention to first select one or the other of a pair of discs and found the pattern of errors predicted. Perceiving these spatial relationships evidently requires selecting and processing objects one by one and is only possible at low object speeds.

Results and Discussion

Viewing a simple scene, say, a line of four colored jellybeans, one has the impression of grasping the colors present and immediately knowing that the red bean is next to the green bean or next to the yellow bean. We asked observers to view an array of colored discs moving at a speed that exceeded the rate at which they could be followed with attention. The observers could still see which colors were present, but they were unable to determine which colors were adjacent. This finding, together with evidence from a cueing experiment, suggests that apprehending which colors are neighbors requires a shift of attention from one object to the other.

Participants were tested in several tasks that used variants of a common display involving six colored discs evenly spaced about a circular path. The array was centered on a fixation point, where observers maintained the center of their gaze (Figure 1). The discs all rotated about the fixation point in the same direction and at the same rate for at least 2.7 s. The ring of six discs was accompanied in some experiments by a second ring of six discs enclosing the first, aligned with it and moving in the same direction and at the same revolution rate.

To measure the speed limit for following a disc with attention, we used a tracking task that resembles the shell games used by conjurers to confound audiences. A magician might cover an object with one of three identical shells and shuffle the shells back and forth while the audience tries to keep their attention on the critical shell. Above a certain object speed, however, attention cannot keep up [4]. To measure this tracking speed limit with our display, we briefly highlighted one disc at the beginning of a track trial (Figure 1). After the discs revolved about fixation for 2.7 s at a particular angular speed, the participants judged which had been highlighted at the beginning. The average angular speed limit (75% threshold; see Figure 2) was 1.4 rps, similar to previous results [4]. The direction of motion of the discs can be judged accurately at speeds much faster than the tracking limit ([4]; see also Supplemental Information available online), suggesting that the tracking limit is imposed by the characteristics of attention rather than by earlier stages of vision [4].

The slow speed limit on attentional tracking is here exploited to investigate the role that attentional tracking plays in the perception of a scene, through comparison of perception below the speed limit to above it. Unlike traditional attentional tests that consume an unknown amount of attention with a secondary task [5, 6] or reduce processing time with brief exposures [6, 7], with the present test, participants can engage their full attentional resources to scrutinize the scene for an extended interval.

We first established that far above the tracking speed limit, participants could still identify the colors present. Each of the six discs was assigned one of three colors drawn randomly without replacement on each trial from a set of six. As in some of our other tests, the sequence of three colors was repeated to yield six discs around the ring (Figure 1, Identify, and Movie S6 provided to illustrate the method at a slow speed). Accuracy for identifying the three displayed colors was high at rates much faster than the tracking speed limit, not falling to 75% correct until speeds exceeded 3.0 rps (Figure 2). The selection by attention afforded by tracking is evidently not needed to perceive the colors.

To determine the maximum speed at which simple spatial relationships could be perceived, we used two tasks. In the within-ring task, participants judged the relative order of the three colors around a ring, in the direction of motion (Figure 1; Movie S4). At the end of a trial, they were presented with the three colors that had been shown, and they clicked on them in order, beginning with any one. In the across-rings task, two concentric rings of six discs were shown. The discs in the outer ring comprised three colors that were distinct from the three colors of the inner ring. Participants were asked to click on any two colors that were aligned with each other. For both the within-ring and across-ring tasks, confidence intervals (Figure 2) and t tests indicate that the speed limits were much slower than the color identification task and were

*Correspondence: alex.holcombe@sydney.edu.au

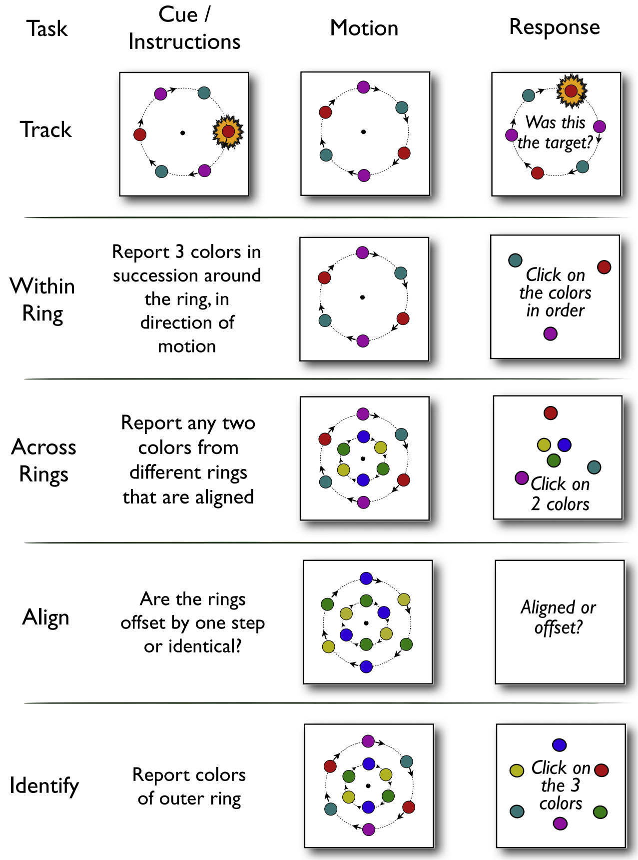

Figure 1. Schematics of the Stimuli and Tasks

The schematics are not to scale, and the "Cue/Instructions" and "Response" column conveys the instructions and response options but does not use the exact same text as the experiments, which can be seen in the supplemental movies.

not significantly different from the 1.4 rps limit on tracking [Figure 2; $t(6) = -0.94$, $p = 0.39$; $t(7) = 0.82$, $p = 0.45$]. Further analyses and an additional experiment in the Supplemental Information address the different chance performance rates of these tasks and find the same pattern of results. So, despite the ability at high object speeds to identify the direction of motion and the colors, participants could only determine which colors were adjacent at slow speeds near or below the tracking limit.

In the spatial relationship tests described above, adjacent colors were always different from each other. To report the colors of two such adjacent discs, it may be that attention must individuate each of the two. In other situations, observers may benefit from preattentive mechanisms that provide information regarding the presence of differences in color or luminance. In particular, differences between neighboring elements provide a signal for the perception of edges and texture boundaries and some aspects of global shape. The underlying perceptual mechanisms are known to efficiently integrate signals from many elements on the basis of brief exposures [8–11], and thus the corresponding percepts likely do not require the selection of elements by attention afforded by tracking. We devised an alignment test that might allow discrimination of spatial relationships using these putatively preattentive mechanisms. Each of the two concentric rings comprised the same three colors in the same order, and participants judged whether the sequences were aligned or instead offset by one color (Figure 1; Movie S5; Movie S6). When they were aligned, neighboring colors in the radial direction were identical. Performance was accurate at rates much faster than the attentive tracking limit and did not fall to 75% accuracy until 2.5 rps (Figure 2).

These results indicate that one can perceive the colors present in a display and their array of locations and potentially discriminate whether adjacent colors are the same or different but not know the identities of different adjacent colors. Attentional tracking is apparently critical to apprehend which colors are neighbors.

Previous findings with stationary displays had already suggested that attentional selection is important for apprehending spatial relationships. Each of the previous lines of evidence, however, is subject to alternative interpretation. For example, search experiments showed that finding elements with a particular spatial relationship is particularly time consuming, whereas finding an individual color is fast [12]. This is expected if spatial relationships are extracted by a process with very limited capacity [13], but high-capacity parallel models are also a viable explanation [1, 14, 15]. The phenomenon of illusory conjunctions [6] more directly dissociates the processing of individual features and their spatial relationship. For instance, after a brief flash of a display containing two colored letters (say, a green 'A' and a red 'B'), on a minority of trials, participants report the correct colors and letters, but in the wrong combination ("green B" and "red A"). Rather than this mistake being due to the absence of attention, however, it might be that conjoining the features requires more signal or longer processing time [16], which might also explain the slow thresholds for pairing color and motion [17, 18].

The results here more firmly support a specific role for attention in judging spatial relationships. The need for attentional selection may arise from a restriction in the number of objects that can be processed at once. An extreme possibility is that the system cannot process even two objects at once, but rather processes them one by one. After a first disc's color is identified, an attention shift to the second would be required.

We found support for this attention-shift theory in the type of errors participants make most frequently in the pairing judgment. Relative to the color the participant reports for the inner ring, the color reported from the outer ring might be correct—the aligned color. Otherwise, it must be the color of the disc one position ahead in the direction of motion or one position behind. We will refer to these errors as "outer-leading" and "outer-trailing" errors, respectively. In the initial pairing experiment, for the three participants who used a version of the program that recorded the type of error, the majority of errors were outer-trailing errors rather than outer-leading (data not shown, but the phenomenon is replicated in the next experiment).

The attention-shift theory can explain the predominance of outer-trailing errors, if some accessory assumptions are made. The first accessory assumption is that participants usually track and identify a disc in the inner ring first. The second is that the planned attention shift toward the aligned disc in the outer ring does not fully take into account the disc's motion, such that by the time it arrives, the targeted disc may have moved on. The trailing disc would soon move into attention's selection window and be the one identified.

An alternative account, and a simpler one, is that the outer ring is perceived as offset relative to the inner ring, in the

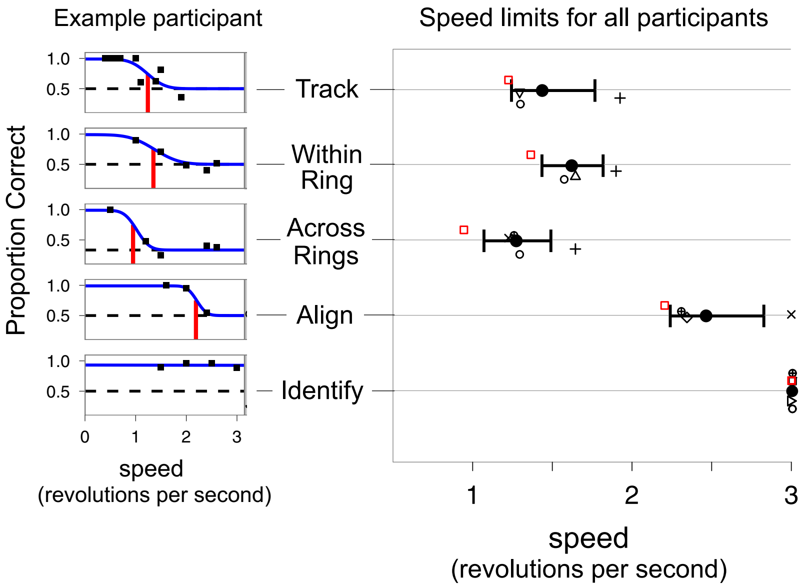

Figure 2. Example Data for One Participant and Speed Limits for All

Left: example data from one participant (author A.O.H.). Right: the speed limits for all observers (different shapes represent different observers, with A.O.H. in red). Each unfilled data symbol shows the 75% threshold for a single observer, estimated in R [26, 27] by fitting cumulative normal distributions via probit regression to proportion of correct responses against speed, after accounting for an assumed 1% lapse rate. For each task, the solid black point shows the mean across observers, and the error bars show the 95% confidence intervals from nonparametric bootstrapping, percentile method [28]. The alignment and identification tasks have faster speed limits than the within-ring, across-ring, or tracking tasks, for which the limits are not significantly different from each other. Thresholds above 3 rps could not be reliably estimated because the motion was spatially undersampled above that rate, a consequence of the 100 Hz refresh rate of the CRT screen. Each observer's identification threshold was in this 3+ rps range.

direction of motion. This account is rejected by the "offset discs" experiment described in the Supplemental Information. In that experiment, we offset the discs of the outer ring relative to the inner ring. A bias toward outer-trailing errors occurred over the whole range of offsets (see Figure S2). For large backward shifts of the outer ring, the outer-leading disc was much nearer the color reported for the inner ring than was the outer-trailing disc, yet still participants usually reported the outer-trailing color. This is expected on the attentional shift account. Participants' attentional shift should target the disc that is closest to being aligned with the selected inner disc, including when it is offset backward. If the targeted disc's motion is not always compensated for, then attention will fall still further behind.

If an attention shift from inner to outer ring is indeed the cause of the error asymmetry, then the error bias should change based on which ring is attended to first. To test this, we manipulated attention with a cue presented at the beginning of the trial, prior to the appearance of the stimulus. The cue was a colored patch. In one block of trials, its color was randomly selected from the discs of the inner ring, and in the other block of trials, its color was randomly selected from the outer ring. Participants judged which color in the other ring was aligned with the cued color. Informal viewing of the display revealed that one can quickly find the cued color, reflecting the effectiveness of feature-based attention for search [19]. On the attention-shift hypothesis, after selecting a disc with the cued color, attention would then shift toward the aligned disc in the opposite ring. When attention is cued to first select a disc in the outer ring, the errors should favor a disc that trails it in the inner ring—an outer-leading error. When the cue instead indicates a color in the inner ring, errors should favor the disc in the outer ring that trails the correct answer—the outer-trailing errors we found before.

The data confirmed the prediction. For trials where the inner ring was cued, outer-trailing errors predominated, and in those where the outer ring was cued, outer-leading errors predominated (Figure 3).

Huang and Pashler have previously provided evidence that the multiple locations occupied in a stationary display can be simultaneously apprehended, whereas the multiple colors present cannot be [20]. Based on one set of studies [7], they even proposed that colors are identified one by one, as we have suggested for the spatial relationship task here. Their proposal may be correct, but their particular results are open to another explanation. They used brief displays with two colored patches that appeared suddenly and were followed by a mask. In support of their theory, when the two colored patches were presented simultaneously, performance was much more accurate for identifying the locations occupied than for identifying the colors present. However, the sudden onset of the color patches would be expected to draw exogenous attention to the patches' locations [21, 22]. Whereas there may not be sufficient time to identify the colors or other features in those locations before the mask terminates featural processing, attention may linger at those locations, providing additional time to encode them.

By using temporally extended displays with objects that can be seen but not tracked, we were able to study the perception of objects without focused attention. Our evidence indicates that spatial relationship judgments require a spatial attention shift. This might be a simple by-product of a limitation on the number of colors that can be identified simultaneously [20]. Alternatively, the shift may play a critical role. Specifically, the direction of the shift together with the temporal order of the colors selected may be used to compute the spatial relationship ([23, 24]; S.L. Franconeri, personal communication).

Experimental Procedures

Eight participants (aged 20–36 including two authors, two females) viewed a 100 Hz CRT screen (Mitsubishi Diamond Pro 2070SB) at 1024 × 768 pixels from a distance of 57 cm. Experiment software was programmed with PsychoPy [25]. All participants had extensive experience fixating during psychophysical experiments. For one participant, the data for the tracking task variation were collected on a Sony GDM F520 CRT screen. Three people participated in the tracking task variation, including two who also participated in the main tracking task.

The stimulus for each task comprised either one or two circular arrays of colored Gaussian-windowed ($\sigma = 0.83°$) filled circles ("discs"), presented against a black screen. The radii of the concentric circular arrays were 2° and 4°, except for two participants for whom 3° and 5° were used. No consistent difference was seen between observers run with radii of 3° and 5° versus those with 2° and 4°, so the results were combined. One observer

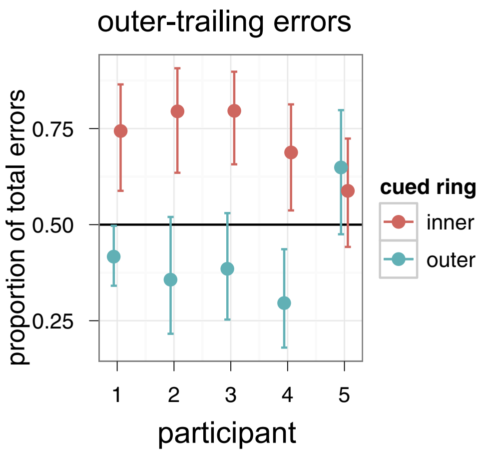

Figure 3. Effect of Cueing on Outer-Trailing Errors

The proportion of errors that were outer-trailing when a disc in the inner ring was cued (red symbols) and when the outer ring was cued (blue symbols), for five participants. The cuing determined which type of error predominated, in the direction predicted by the attention shift account. Error bars show 95% confidence intervals [29].

was run in both conditions as a further check that there was no significant difference, which also confirmed a previous finding of no significant effect of radius on tracking speed limit [4]. Six discs were equally spaced in a circular array and for some tasks consisted of two successive identical sequences of three colors (ABCABC; see Figure 1). These three colors were drawn from a set of six—red (Commission Internationale de l'Eclairage x = 0.61, y = 0.36, 18 cd/m^2), green (x = 0.29, y = 0.58, 25 cd/m^2), blue (x = 0.15, y = 0.07, 15 cd/m^2), yellow (x = 0.40, y = 0.49, 15 cd/m^2), fuchsia (x = 0.27, y = 0.14, 18 cd/m^2), and cyan (x = 0.21, y = 0.28, 25 cd/m^2). The luminances and color coordinates were slightly different for the participant who used the Sony screen.

The duration of the stimulus was 3 s, with the exception of the tracking task that included an additional 400 ms cueing period and did not include a gradual shift to gray at the end. The trial began with a fixation point, following which the array of discs (which varied with task; see below) appeared. To prevent participants from perceiving the colors and their relative locations from the discs' initial stationary state, at the beginning of the trial the discs were gray (x = 0.28, y = 0.28, 32 cd/m^2). They revolved about fixation at a constant rate. The initial angle or phase of the circular array was set randomly on each trial. Over a period of 300 ms, the discs smoothly increased in saturation (linearly through RGB space) to become the appropriate colors (which varied according to task; see below). Subsequently, they continued to revolve for 2 s before smoothly shifting back to gray over a 700 ms period.

The discs then disappeared (except in the tracking task; see below) and the response screen was presented until the participant made her response, for which there was no time limit or pressure. A participant's first session with a task was preceded by between 10 and 40 practice trials at a variety of speeds. Subsequent blocks were preceded by fewer than 10 practice trials. A block of trials consisted of between 70 and 160 trials. The method of constant stimuli was used with several different speeds presented in random counterbalanced order. Direction of motion was crossed with speed and also counterbalanced.

Track Task, Main

All discs began with their appropriate color, except the randomly chosen target disc to be tracked, which was gray (32 cd/m^2). A single circular array of six discs was shown at a radius of 2° (Figure 1). While all the discs revolved about the fixation point, the target disc remained gray for 400 ms, after which it gradually changed over 300 ms into its final color (linearly through RGB space). After the discs continued moving for an additional 2.7 s, one disc was indicated by suddenly changing to gray while it and the rest of the array continued to revolve. The observer judged whether the indicated disc was the one that had been indicated at the beginning. In half of the trials, it was indeed the precued disc, and in the remaining trials, it was the other disc that had the same color as the precued disc, which was always on the opposite side of the circular array. The discs continued revolving until the participant made her response by pressing one of two keys. Guessing would result in 50% correct. See also Movie S1.

Track Task, Variation

This display differed in that all of the discs in a ring were the same color, and participants had to click on which of the six had been the target. Two circular arrays of six discs were presented, in the same positions as for the other tasks. On each trial, one color was randomly chosen and used for all six discs in the outer ring, and another color randomly chosen from the remaining five colors was used for all six discs in the inner ring. A randomly chosen disc in the outer ring together with the disc aligned with it in the inner ring were designated as targets by initially appearing in gray. They subsequently changed gradually to the color of the other discs in their ring, following the same temporal parameters as in the main track task. After the discs revolved about fixation in the same manner as the main track task (with both rings moving in the same direction and always aligned), the participant was required to indicate which disc was a target by moving a mouse pointer and clicking on it. Either the inner disc or the outer disc could be clicked on, but because they were always aligned, only one needed to be tracked. Guessing would result in 1/6 or 17% correct. The results were similar to the main tracking task and are described in the Supplemental Information. See also Movie S2.

Across-Ring Task

Two concentric circular arrays of six discs were presented, with three colors randomly chosen for the outer array and the remaining three of the six possible used for the inner array. Observers determined, for any disc they chose, which disc was aligned with it in the other array. After the discs finished revolving about fixation, the response screen showed the three colors of each ring arrayed in random order. Participants used the mouse to select two discs they judged to have been aligned. Guessing would result in 33% correct. See also Movie S3.

Within-Ring Task

A single array of six discs (three colors in a random sequence repeated twice) was presented, with a radius of 2°. The response screen showed the three colors of the ring arrayed in random order, and participants used the mouse to select three in order. Observers were instructed to report the three colors in the order they appeared about the circular array, in the direction of motion. As there are only two possible relative orders in a circular sequence of three, guessing would result in 50% correct. See also Movie S4.

Alignment Task

The same display was used as in the across-ring task, except that only three colors were randomly drawn from the set of six described above, and these were used in both the outer and inner rings of discs. The three colors were arrayed in the same order in both rings. In half of the trials, they were exactly aligned, such that any line passing through a disc and fixation connected only discs of the same color. In the other half of the trials, the two rings were offset by one disc. Hence, although each disc in the inner ring was aligned with another in the outer ring, the aligned discs were not the same color in the nonaligned trials. Participants judged whether common colors were aligned or not. Guessing would result in 50% correct. See also Movie S5 and Movie S6.

Identify Task

Participants reported the three colors presented. The stimulus configuration was that of the two concentric arrays used in the pairing task. However, rather than a distinct triplet of colors used in the inner and outer arrays as was used in the pairing task, here the colors in the two arrays were chosen independently from each other, so that the identity of the colors in one array provided no information about the other array. The response screen showed the six possible colors, and the participant used the mouse to select the three that had been presented in a particular array (inner or outer). In some blocks of trials, participants were instructed to report the colors in the inner circular array; in other cases, those of the outer. Guessing would result in 50% correct. See also Movie S7.

Supplemental Information

Supplemental Information includes two figures, Supplemental Experimental Procedures, and seven movies and can be found with this article online at doi:10.1016/j.cub.2011.05.031.

Acknowledgments

We thank Patrick Cavanagh, whose ideas partially inspired the project and in whose graduate seminar some of this work was conceived. His National Institutes of Health grant EY009258 supported M.V.-P. Discussions with Steven Franconeri were very helpful, as was his and his colleagues' unpublished manuscript on the attention-shift hypothesis. Jeremy Wolfe provided helpful comments. D.L. was supported by an Australian Research Council Discovery Project awarded to A.O.H. and a Beautriu de Pinós fellowship from the Generalitat de Catalunya (Spain). A.O.H. was supported by an Australian Research Council Future Fellowship grant.

Received: March 23, 2011
Revised: April 28, 2011
Accepted: May 15, 2011

References

1. Palmer, J. (1994). Set-size effects in visual search: The effect of attention is independent of the stimulus for simple tasks. Vision Res. *34*, 1703–1721.
2. Wolfe, J.M., and Bennett, S.C. (1997). Preattentive object files: Shapeless bundles of basic features. Vision Res. *37*, 25–43.
3. Lee, D.K., Koch, C., and Braun, J. (1999). Attentional capacity is undifferentiated: Concurrent discrimination of form, color, and motion. Percept. Psychophys. *61*, 1241–1255.
4. Verstraten, F.A.J., Cavanagh, P., and Labianca, A.T. (2000). Limits of attentive tracking reveal temporal properties of attention. Vision Res. *40*, 3651–3664.
5. Pastukhov, A., Fischer, L., and Braun, J. (2009). Visual attention is a single, integrated resource. Vision Res. *49*, 1166–1173.
6. Treisman, A., and Schmidt, H. (1982). Illusory conjunctions in the perception of objects. Cognit. Psychol. *14*, 107–141.
7. Huang, L., Treisman, A., and Pashler, H. (2007). Characterizing the limits of human visual awareness. Science *317*, 823–825.
8. Motoyoshi, I., and Nishida, S. (2001). Temporal resolution of orientation-based texture segregation. Vision Res. *41*, 2089–2105.
9. Wilson, H.R., and Wilkinson, F. (1998). Detection of global structure in Glass patterns: Implications for form vision. Vision Res. *38*, 2933–2947.
10. Clifford, C.W.G., Holcombe, A.O., and Pearson, J. (2004). Rapid global form binding with loss of associated colors. J. Vis. *4*, 1090–1101.
11. Ramachandran, V.S., and Rogers-Ramachandran, D.C. (1991). Phantom contours: A new class of visual patterns that selectively activates the magnocellular pathway in man. Bull. Psychon. Soc. *29*, 391–394.
12. Wolfe, J.M. (1998). What can 1 million trials tell us about visual search? Psychol. Sci. *9*, 33–39.
13. Treisman, A.M., and Gelade, G. (1980). A feature-integration theory of attention. Cognit. Psychol. *12*, 97–136.
14. Thornton, T.L., and Gilden, D.L. (2007). Parallel and serial processes in visual search. Psychol. Rev. *114*, 71–103.
15. McElree, B., and Carrasco, M. (1999). The temporal dynamics of visual search: Evidence for parallel processing in feature and conjunction searches. J. Exp. Psychol. Hum. Percept. Perform. *25*, 1517–1539.
16. Tsal, Y., Meiran, N., and Lavie, N. (1994). The role of attention in illusory conjunctions. Percept. Psychophys. *55*, 350–358.
17. Arnold, D.H. (2005). Perceptual pairing of colour and motion. Vision Res. *45*, 3015–3026.
18. Holcombe, A.O. (2009). Seeing slow and seeing fast: Two limits on perception. Trends Cogn. Sci. *13*, 216–221.
19. Shih, S.-I., and Sperling, G. (1996). Is there feature-based attentional selection in visual search? J. Exp. Psychol. Hum. Percept. Perform. *22*, 758–779.
20. Huang, L., and Pashler, H. (2007). A Boolean map theory of visual attention. Psychol. Rev. *114*, 599–631.
21. Nakayama, K., and Mackeben, M. (1989). Sustained and transient components of focal visual attention. Vision Res. *29*, 1631–1647.
22. Solomon, J.A. (2004). The effect of spatial cues on visual sensitivity. Vision Res. *44*, 1209–1216.
23. Cavanagh, P., Hunt, A.R., Afraz, A., and Rolfs, M. (2010). Visual stability based on remapping of attention pointers. Trends Cogn. Sci. *14*, 147–153.
24. Franconeri, S., and Handy, T. (2007). Rapid shifts of attention between two objects during spatial relationship judgments. J. Vis. *7*, 582.
25. Peirce, J.W. (2007). PsychoPy—Psychophysics software in Python. J. Neurosci. Methods *162*, 8–13.
26. R Development Core Team. (2010). R: A Language and Environment for Statistical Computing (Vienna: R Foundation for Statistical Computing). http://www.R-project.org/.
27. Wickham, H. (2009). Ggplot2: Elegant Graphics for Data Analysis (New York: Springer).
28. Efron, B., and Tibshirani, R. (1994). An Introduction to the Bootstrap (Virginia Beach, VA: Chapman & Hall/CRC).
29. Clopper, C.J., and Pearson, E.S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika *26*, 404–413.