Nature, 2000, Nov 9

Tracking an object through feature-space

Erik Blaser, Zenon Pylyshyn, & Alex O. Holcombe*

Rutgers Center for Cognitive Science
Rutgers University
Piscataway, NJ 08854, USA

**Department of Psychology*
Harvard University
Cambridge, MA 02138, USA

Correspondence and requests for materials should be addressed to E.B. (e-mail: blaser@ruccs.rutgers.edu).

This research was supported in part by NIH grant 1R01-MH60924 to ZP.

Nature, 2000, Nov 9

By attending to a moving item, one can keep track of it among distractors solely on the basis of its spatial location, even when the distractors share the same features. We show that one can also track through *feature-space*: one can keep track of an item solely on the basis of its trajectory along colour, orientation, and spatial frequency dimensions, even when a distractor shares the same spatial location. This demonstration, coupled with our finding of competition for attentional resources between superimposed items but not within them, implies the operation of object-based attention and demonstrates that distinct *featuro*-temporal trajectories are sufficient for objecthood and attentional tracking.

Even with the eyes fixed, observers are able to attentively track a cued target, or targets, even in the presence of identical, but irrelevant distractor items (ref. 1). In previous studies, a target item was distinguished from distractors by its location in the display. But what if the target could not be distinguished by its location, but only by its position in a "feature-space" spanned by, say, colour, orientation and spatial frequency? What if this target item were spatially superimposed with a competing distractor and both were set into motion, drifting through this feature-space while remaining stationary on the screen? Would observers be able to attentionally track such a target over time? Here we show that observers are indeed able to track a target solely on the basis of its position in a feature-space. Given this, the definition of "visual object" (ref. 2) must be extended to include not only items with well-defined spatio-temporal trajectories (ref. 3), but also those with well-defined featuro-temporal trajectories. Moreover, the ability of observers to attentively track an item through feature-space bears directly on the debate between "location-", "feature-", and "object-based" theories of attention, that is, whether attention selects locations in space (ref. 4, 5); feature dimensions such as colour, size, or shape (ref. 6, 7); or particular visual "objects" composed of constellations of visual features. The ability to track through feature-space shows, consistent with a growing body of psychophysical (ref. 8-11) and neurophysiological (ref. 12) evidence, that attention can indeed select entire visual objects.

In our investigation, observers tracked and made judgments about a complex time-varying stimulus. The basic stimulus was a pair of gabor patches that randomly, yet smoothly, drifted in a feature-space defined by orientation, spatial frequency, and colour. The trajectory for each gabor, and along each of the feature dimensions, was random and independent, so the gabors frequently changed direction and speed along any given dimension (they did, however, drift with some "inertia" which made them more likely to continue changing in the direction they were going). Consistent with their trajectory along these dimensions, the gabors smoothly rotated, spinning sometimes clockwise and sometimes counter-clockwise; changed gradually between a few broad bars to many thin bars; and altered their saturation, changing smoothly in appearance between a patch defined by gray and black bars to a patch defined by red and black bars, through all the intervening saturation levels of red (Fig. 1a).

With a single item, location and "objecthood" are confounded (ref. 13). We unconfounded location and objecthood in our experiments by completely superimposing these two gabors such that there was no spatial distinction to support selection by location-based attention. The gabors were superimposed by temporally interleaving their images at a rate of 117 Hz (nearly 60 Hz per item, with one gabor shown in the even video frames, and the other in the odd frames). However, the gabors appeared to be simultaneously present and transparently layered on one another. Importantly, individual parts of each of the gabors were beyond the spatial resolution of attention (ref. 5, 14): the spatial frequency of a 1 degree diameter gabor ranged dynamically between 1.5 to 8 cycles per degree (consequently, the width of a single bar ranged from 20 and 4 arcmin). Without this restriction, previous studies using superimposed figures (ref. 8) which produced results consistent with object-based models may be influenced by a strategy where location-based attention is switched between parts of the compound stimulus, thereby mimicking object-based attention.

In Experiment 1, observers were shown a display which consisted of the two superimposed gabors centered on fixation (Fig. 1b). During the 10 second tracking interval, both gabors remained stationary at fixation, though they drifted through colour-orientation-frequency space. The gabors began at different places in this space, and observers were instructed to track the "target" gabor, which was identified by being initially oriented at 45 degrees. Since the motion of each gabor along each of the feature dimensions was random and independent, the gabors frequently "passed" each other along any given dimension (Fig. 1c). (This is a critical property of the stimulus, because it ensures that feature-based attention could not simply pick out the target on the basis of some constant featural difference. Unless this is done, results which are consistent with object-based attention are subject to criticisms as artifacts of feature-based attention. For instance, this is of concern in studies where two "objects" always have, say, different motion directions throughout the entire trial (ref. 9). Here one could argue that though location-based attention may be foiled by the superimposition of two such items, feature-based attention could still pick out a cued item simply on the basis of this constant featural difference.) At the end of the tracking interval, both items stopped changing and observers were asked to pick out the gabor that corresponded to the target. On average, the four observers (naive: JT, DA, DK; expert: EB) successfully picked the target item in 90% of the trials.

How were observers able to accomplish this tracking through feature-space? Revealingly, observers reported that the two superimposed gabors perceptually segregated from one another. The segregation of superimposed stimuli has been observed previously in studies of motion transparency (ref. 15), and attention-based motion perception (ref. 6). Additionally, judging from informal observations of our stimuli, the simultaneous "motion" along the orientation, colour, and spatial frequency dimensions results in much stronger transparency than that which results from motion along any one, or pair, of these dimensions alone. Beyond the perception of transparency, observers reported that attending to a particular gabor resulted in its introspective enhancement, in a manner not unlike figure-ground segmentation. The attended gabor became figure and the distractor gabor receded into an ignored ground (ref. 16). This is in itself suggestive evidence for the involvement of object-based attention (and is a phenomenon which is unobservable in studies which use briefly presented stimuli instead of time-varying and extended stimuli).

For each of the attention models under consideration here, this is a very different stimulus. Feature-based models treat the stimulus as a mix of colours, orientations, and spatial frequencies, each of which may be selectively enhanced by attention; location-based models treat the stimulus as a unitary stream of information, the entirety of which may be enhanced (ref. 17); and object-based attention models treat the stimulus as *two* unitary streams of information, each corresponding to one of the superimposed gabor patches, and each a legitimate target for visual attention. Moreover, object-based models predict that when one attempts to attend to a particular region or feature of one of the gabors in our stimulus, one is actually, by default, attending to the object as a whole; as a consequence, processing is enhanced for all of the constituent features of the gabor, irrespective of type or number (ref. 10). This leads to two predictions which can tested psychophysically to distinguish object- from location- and feature-based attention. First, in contrast to location-based models, object-based models predict that selective enhancement of an object can take place even if there is another object which is spatially superimposed. Second, in contrast to feature-based models, selective attention to any feature of an object should also enhance processing of its other, irrelevant feature dimensions.

In Experiment 2, we tested these predictions by having observers perform two concurrent detection tasks using these same two superimposed gabors. At some random time in a 5 second trial, both gabors exhibited a slight discontinuity in their motion along all 3 dimensions (colour, spatial frequency, and orientation) simultaneously; that is, there was a slight "jump" in the trajectory of each gabor (Fig. 2). The direction of the jumps was chosen randomly and independently for each gabor and dimension, and the sizes of the jumps were fixed at values corresponding to 75% correct thresholds determined from baseline psychometric functions of jump-size for each dimension and observer. In any given block of trials, observers were instructed to concurrently attend to a pair of dimensions and to report the direction of the jump (e.g., "left" versus "right", "more red" versus "less red") for both. The 3 conditions formed by all possible pairs of features were further blocked by attention allocation instructions. Observers were either instructed to attended equally to the given pair of dimensions, or to attempt to primarily attend to one of the two. For instance, in a particular block of trials, observers were asked to report the direction of orientation and colour jumps while attempting to devote most of their attention to orientation and relatively little attention to colour. Single-task control conditions were also run, where observers only had to attend to, and make judgments about, one feature dimension.

By manipulating attention instructions in this way, Attention Operating Characteristics (AOC's) (ref. 18) were measured. The area under these AOC's measures the extent to which concurrent tasks, here the two jump-direction judgments, compete for processing resources. When two tasks do not compete for resources at all, performance on one of the tasks should not suffer when the other task is being performed concurrently (like walking and chewing gum). When two tasks do compete for resources performance on one of the tasks will suffer when the other task is being performed concurrently (like juggling and sign-language). We quantify this competition for resources with a "relative area" measure which is the ratio of the area under the AOC to the theoretical maximum area that would be achieved if there were no competition for resources between the two tasks at all. Low relative areas indicate competition for resources, while high relative areas indicate little competition. Since competition occurs when an attention mechanism is forced to divide its resources between its legitimate attentional targets, it is possible determine which type of attention is mediating performance by observing when competition occurs.

The critical comparison in this experiment is between conditions where observers made the pair of jump-direction judgments within a common gabor, and conditions where the same two judgments were divided between the two gabors (e.g., report the colour jump-direction of one gabor, and the orientation jump-direction of the other). In accordance with the above predictions, since object-based attention enhances processing of all the features of an attended object, there should be little or no competition for resources *if the concurrent judgments are made with respect to a common gabor*. However, if concurrent judgments are divided between the two gabors, significant competition should result, since now attentional resources must be divided between the objects. This is exactly the pattern of results we found: Observers had an average relative area of 94% when attending to, and making judgments within one of the two objects, whereas the area dropped to 64% when observers attempted to divide their attention to make the judgments between the two objects (Fig. 3a, 3b). These results cannot be accounted for by either location- or feature-based models of attention, and support the predictions of object-based models.

Experiment 1 showed that observers are capable of tracking a single object in spite of a spatially superimposed distractor. But what is the upper limit of this ability? Are observers capable of tracking two objects? Three objects? In the between-object conditions of Experiment 2, observers had both an instruction and a task which encouraged them to attend to both objects simultaneously. If observers were truly able to "share" attention between the two objects, that is, attend to and track both simultaneously, then the correctness of responses on the two judgments should show statistical independence. That is, whether or not observers were correct on the jump-direction judgment with respect to "object 1" should not significantly influence their correctness with respect to "object 2". Alternatively, if observers were unable to attend to both simultaneously, and instead "switched" attention-- sometimes attending to and tracking one object, and sometimes attending to and tracking the other-- then the correctness of responses should show a negative correlation (i.e., if observers were correct on object 1, then they should be less likely to be correct on object 2, and visa versa) (ref. 18). Subjecting these data to a Chi-square analysis rejected the null hypothesis of statistical independence, and therefore of pure sharing of attention. Additionally, the data show a clear negative correlation, suggesting that observers were switching their attention between the two objects. The result of this analysis, taken together with observers' reports of an inability to attend to both objects simultaneously, suggests that the upper limit on tracking through feature-space is a single object.

We have shown that observers are able to track a target in spite of a spatially superimposed distractor, in a paradigm requiring observers to monitor the target's changing position in a feature-space spanned by the dimensions of colour, orientation and spatial frequency. The ability to track through feature-space is likely mediated by a system which, though capable of switching attention between objects, can only track one object at a time. The ability to track through feature-space, coupled with our finding of competition for attentional resources between superimposed items, but not within them, not only implies the operation of object-based attention, but also shows that distinct "visual objects" need not have distinct spatio-temporal trajectories-- rather, distinct *featuro*-temporal trajectories are sufficient for objecthood and attentional tracking.

REFERENCES AND NOTES

1. Z. W. Pylyshyn, R. W. Storm, *Spatial Vision* **3**, 179-197 (1988); S. Yantis, *Cognit Psychol* **24**, 295-340 (1992).
2. A. Treisman, R. Paterson, *J Exp Psychol Hum Percept Perform* **10**, 12-31 (1984).
3. C. von Hofsten, E. S. Spelke, *J Exp Psychol Gen* **114**, 198-212 (1985); J. Stiles-Davis, *J Exp Psychol Gen* **115**, 96-7 (1986); E. S. Spelke, C. von Hofsten, *J Exp Psychol Gen* **115**, 98-100 (1986); C. von Hofsten, P. Vishton, E. S. Spelke, Q. Feng, K. Rosander, *Cognition* **67**, 255-85 (1998).
4. M. I. Posner, *Q J Exp Psychol* **32**, 3-25 (1980); D. Sagi, B. Julesz, *Nature* **321**, 693-695 (1986).
5. C. W. Eriksen, J. E. Hoffman, *Percept Psychophys* **12**, 201-204 (1972).
6. P. Cavanagh, *Science* **257**, 1563-5 (1992).
7. J. M. Wolfe, K. R. Cave, S. L. Franzel, *J Exp Psychol Hum Percept Perform* **15**, 419-33 (1989); M. Corbetta, F. M. Miezin, S. Dobmeyer, G. L. Shulman, S. E. Petersen, *Science* **248**, 1556-9 (1990).
8. J. Duncan, *J Exp Psychol Gen* **113**, 501-17 (1984).
9. M. Valdes-Sosa, A. Cobo, T. Pinilla, *Cognition* **66**, B13-B23 (1998).
10. S. J. Luck, E. K. Vogel, *Nature* **390**, 279-81 (1997).
11. D. Kahneman, A. Treisman, B. J. Gibbs, *Cognit Psychol* **24**, 175-219 (1992); G. C. Baylis, J. Driver, *J Exp Psychol Hum Percept Perform* **19**, 451-70 (1993); A. Treisman, *Philos Trans R Soc Lond B Biol Sci* **353**, 1295-306 (1998).
12. J. Driver, G. C. Baylis, R. D. Rafal, *Nature* **360**, 73-5 (1992); J. Driver, G. C. Baylis, S. J. Goodrich, R. D. Rafal, *Neuropsychologia* **32**, 1353-65 (1994); M. Valdes-Sosa, M. A. Bobes, V. Rodriguez, T. Pinilla, *J Cogn Neurosci* **10**, 137-51 (1998); K. M. O'Craven, P. E. Downing, N. Kanwisher, *Nature* **401**, 584-7 (1999).
13. J. S. Lappin, *J Exp Psychol* **75**, 321-8 (1967); A. Bonnel, W. Prinzmetal, *Percept Psychophys* **60**, 113-124 (1998).
14. S. He, P. Cavanagh, J. Intriligator, *Nature* **383**, 334-7 (1996); E. Blaser, G. Sperling, Z. L. Lu, *Proc Natl Acad Sci U S A* **96**, 11681-6 (1999).
15. N. Qian, R. A. Andersen, E. H. Adelson, *J Neurosci* **14**, 7357-66 (1994).
16. U. Neisser, R. Becklen, *Cognit Psychol* **7**, 480-494 (1975); I. Rock, D. Gutman, *J Exp Psychol Hum Percept Perform* **7**, 275-85 (1981).
17. S. I. Shih, G. Sperling, *J Exp Psychol Hum Percept Perform* **22**, 758-79 (1996).
18. G. Sperling, M. J. Melchner, *Science* **202**, 315-8 (1978); G. Sperling, in *Varieties of Attention* R. Parasuraman, D. R. Davies, Eds. (Academic Press, New York, 1984) pp. 103-181; D. K. Lee, C. Koch, J. Braun, *Percept Psychophys* **61**, 1241-55 (1999).

FIGURE CAPTIONS

Figure 1 (a) Snapshots taken every 250 msec from a 2.5 second sequence of two superimposed gabor patches drifting in a feature-space spanned by the dimensions of colour, orientation, and spatial frequency. In Experiment 1, observers were cued to track one of these gabors. After a 10 second sequence, observers indicated which of the gabors corresponded to the cued target (by using the small number-labels shown here in the last snapshot). **(b)** Schematic of display, illustrating that the two gabors were stationary in space, and superimposed on fixation. The compound stimulus subtended 1 degree of visual angle. **(c)** Schematic of the two gabors drifting in feature-space. Gabors drifted independently and randomly (though with some "inertia" which made a gabor more likely to drift in the direction it was going). Gabors frequently changed direction and speed and "passed" each other in this space.

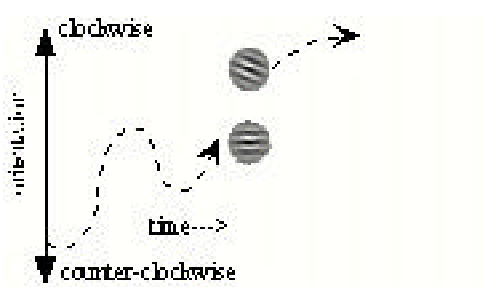
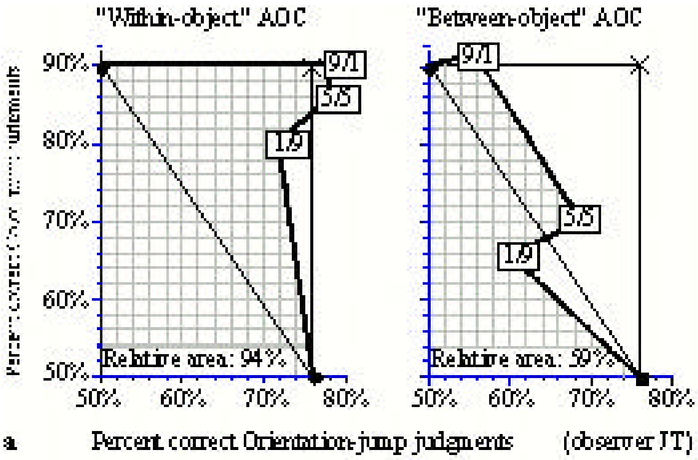


Figure 2 Schematic of a clockwise "jump" along the orientation dimension. By re-labeling the left axis, this same figure can represent the two potential jump-directions ("clockwise" versus "counter-clockwise", "more red" versus "less red",

and "higher frequency" versus "lower frequency") for any of the dimensions (i.e., orientation, colour, and spatial frequency, respectively) used in Experiment 2.



a

Condition	"Within"	"Between"	Observer
Spatial	94%	59%	JT
frequency	87%	50%	DA
&	98%	68%	DK
Color	100%	60%	EB
Color	93%	80%	JT
&	100%	69%	DA
Orientation	78%	69%	DK
	95%	59%	EB
Orientation	93%	60%	JT
&	92%	71%	DA
Spatial	94%	70%	DK
frequency	100%	56%	EB

b

Figure 3 (a) Two example Attention Operating Characteristics from Experiment 2 showing performance in a "Within-object" judgment condition and the corresponding "Between-object" condition, respectively. The attention allocation instructions that were given to the observer are indicated on the data points, with "9/1" referring to 90% dimension 1 (here, colour) attention and 10% dimension 2 (orientation) attention; "5/5" referring to 50% attention to colour, 50% to orientation; and "1/9" referring to 10% attention to colour, 90% to orientation. Solid dots on the axes represent performance in the single-task control conditions. The "X" symbol is the AOC's "independence point", given by the intersection of the coordinates of these two control conditions. The area subtended by connecting the control condition points to the independence point is the theoretical maximum area, given no competition for resources during the two concurrent jump-direction judgment tasks. The oblique line connecting the two control conditions marks-off the theoretical minimum area, given a total trade-off of resources between the two tasks. Relative areas are calculated by taking the ratio of the area under the observer's AOC (the cross-hatched region) to the theoretical maximum area. **(b)** Table showing relative area measures for "Within-" and "Between-object" conditions for pairs of dimensions, and each observer (naive: JT, DA, DK; expert: EB). The fact that the relative areas for "Within" conditions are high, and larger than those for "Between" conditions, indicates that two concurrent judgments made within a single gabor are accomplished without significant competition for attentional resources, while judgments split between the two gabors compete significantly. This pattern of results is inconsistent with location- and feature-based attention models, but is indicative of object-based attention.