Rapid communication

A simple saliency model predicts a number of motion popout phenomena

Ruth Rosenholtz

Xerox PARC, 3333 Coyote Hill Road, Palo Alto, CA 94304, USA

Abstract

Visual search for a moving target among stationary distractors is more efficient than searching for a stationary target among moving distractors, and searching for a fast target among slow distractors is more efficient than vice versa. This indicates that the ease of search for a target with a particular motion is not determined simply by the difference between target and distractor velocities. We suggest a simple model for predicting ease of search for a unique motion, based upon a quantitative measure of target saliency. Essentially, search will be easier the more the target motion deviates from the general pattern of velocities in the scene. Our model predicts a number of well-known motion search phenomena, and suggests that one control for target saliency as well as target discriminability when drawing conclusions about visual system mechanisms from search experiments. © 1999 Published by Elsevier Science Ltd. All rights reserved.

Keywords: Popout phenomena; Stationary target; Distractor velocity; Motion search; Linear separability

1. Introduction

The term popout is often used to describe a bottom-up phenomenon in which a part of a scene or display seems to draw our attention, or pop out at us. One typically attempts to judge the saliency, or degree to which a scene element pops out, by studying the efficiency of search for that item. The reasoning is that if an item draws our attention, search for that item should proceed more efficiently than search for an item that does not draw our attention. Search efficiency is typically judged by looking at the slope of the plot of reaction time (RT) vs. number of items in the display (set size). We can use criteria such as that suggested in (Wolfe, 1996), in which slopes of less than about 10 ms/item indicate fairly efficient search, and slopes greater than 20 ms/items mark inefficient search.

Ecologically, it makes sense for a popout mechanism to exist. It may often be necessary for survival to notice objects that, for instance, move differently from their surroundings. We should notice such objects even if we are not explicitly looking for them, since they may be predator, or prey, or may be about to collide with the observer.

One naïve hypothesis might be that search efficiency would be determined by the distance in feature space between the feature value for the target and for the most similar distractors. For example, for motion search, one might expect search efficiency to be determined by the difference between the target velocity and the velocity of the distractors that moved most like the target. However, for motion search (as for other search modalities), a number of search asymmetries show that the story is not this simple: search for a moving target among stationary distractors is easier than search for a stationary target among moving distractors (Dick, Ullman & Sagi, 1987; Dick, 1989; Klempen, Shulman, Royden & Wolfe, 1998); search for a fast target among slow distractors is more efficient than search for a slow target among fast distractors (Ivry & Cohen, 1992); and adding variability in speed when searching for a unique motion direction has little effect, while adding variabil-

E-mail address: rruth@parc.xerox.com (R. Rosenholtz).

ity in direction when searching for a unique speed makes the search task more difficult (Driver, McLeod & Dienes, 1992). Currently there has been no coherent explanation of all of these results; each new experiment has generated a new explanation of the new effect.

We present a model for the bottom-up mechanism behind motion popout that explains all of these motion search results. We start by representing the motion of each display element as a point in velocity space, (v_x, v_y). From the distribution of the motions present in the display, we compute the mean and covariance of the distractor[1] motions, μ and Σ, respectively[2]. We then define target saliency as the Mahalanobis distance, Δ, between the target velocity, v, and the mean of the distractor distribution, where

$$\Delta^2 = (v - \mu)^T \sum{}^{-1} (v - \mu)$$

Essentially, we are using as the measure of target saliency the number of standard deviations between the target velocity and the mean distractor velocity. In the simplest version of the model, the more salient the target, the easier the search.

Adding internal noise to either the distractor or target observations is equivalent to convolving the distractor distribution with the noise distribution, which for the case of Gaussian noise is merely equivalent to adding additional noise terms to the covariance, Σ. For the predictions shown in this paper, we arbitrarily added isotropic, normally-distributed noise with standard deviation 0.2 deg/s, to observations of both target and distractor velocities. We predict the same asymmetries for zero internal noise, and for internal noise standard deviations at least as large as 0.45 deg/s. In a more complicated version of this model, it might be appropriate to have the magnitude of the noise depend upon the velocity. This would account for Weber's law behavior, in which motions are less discriminable at a higher velocity. One might also want different noise in speed than in direction, as opposed to the isotropic noise used here.

The target is more likely to pop out the greater the distance, Δ, between its velocity and the mean of the distractor distribution. In many cases, one can immedi-

ately tell from the representation of the stimuli in velocity space whether the model predicts efficient or inefficient search. In our plots, we represent the mean and covariance of the distractors by the 1σ covariance ellipse, centered at the mean distractor velocity. When the target falls within this ellipse, we predict inefficient search. For more subtle cases we calculate the saliency, Δ, to clarify the prediction.

The notion of target saliency is reminiscent of the bottom-up portion of Wolfe's *activation map* (Wolfe, 1994), but his computation of activation differs from our measure of saliency. Our measure of saliency quantifies observations (Nothdurft, 1937; Duncan & Humphreys, 1989) that search becomes easier when target-distractor difference is large relative to the variability in the distractors. Finally, this model is similar to work in color search which suggests that search is easy if the target is linearly separable from the distractors in color space (D'Zmura, 1991; Bauer, Jolicoeur & Cowan, 1996). However, our continuous measure of saliency is more powerful than the notion of linear separability. In its current form, the linear separability model gives only a binary judgment of whether or not search will be easy; in (Bauer, Jolicoeur & Cowan, 1996) the authors demonstrate the increased ease of search with increased distance between the target and the line of linear separability, but their model does not quantify this. In addition, our model differs from the linear separability model in that it cares about the mean and covariance of the distractors, and not merely the location of the line separating the target from the distractors. We discuss this difference in more detail below, in Section 2.1.

2. Predictions of motion search results

One of the classic motion popout asymmetries is that a moving target may be detected among stationary distractors much more easily (Dick, Ullman & Sagi, 1987; Klempen et al., 1998) than a stationary target may be detected among moving distractors (Dick, 1989; Klempen et al., 1998). It should be clear why the former case is easy from Fig. 1A. The distractors are all clustered about the origin, and for any reasonable target speed, the target lies well outside the 1σ covariance ellipse, and therefore we predict that the moving target pops out.

In classic demonstrations of search for a stationary target among moving distractors, the distractors move randomly, or, as shown in Fig. 1B, in opposite directions with random phase. In either of these cases, the target lies right at the mean of the distractor distribution, and our model predicts that search would be difficult. This brings up the obvious question of what would happen if the distractors moved coherently in

---

[1] We suggest that, in practise, the visual system computes the mean and covariance of *all* of the motions present in the display, thus requiring no advance knowledge of which element is the target. For displays with a reasonably large number of distractors and at most one target, the target velocity has little effect on this computation, and it is as if we used only the distractor velocities. Similarly, saliency would be computed for both targets and distractors.

[2] Alternatively, one may represent motion as orientation in *xyt* space. This representation of motion was suggested by Adelson & Bergen (1985), van Santen and Sperling (1984) and Watson and Ahumada (1985).
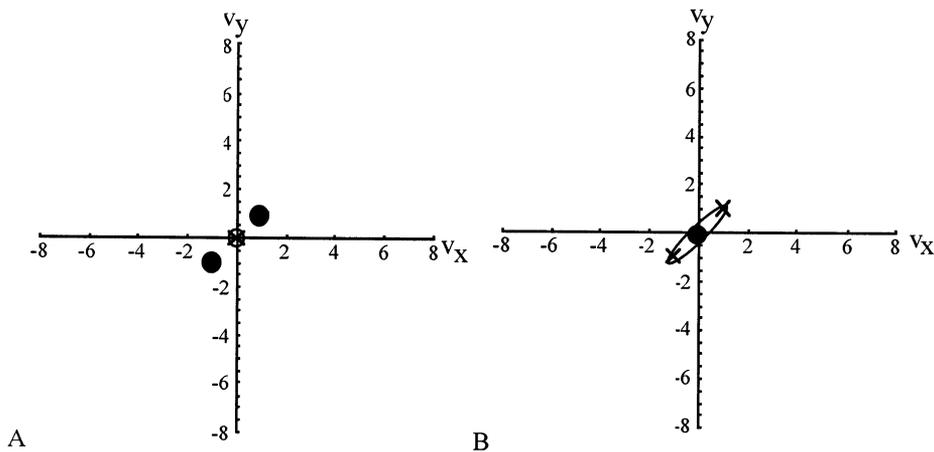
Fig. 1. (A) Search for a moving target (●) among stationary distractors (X). It is assumed that the target is oscillating, and thus at any given instant, it moves either up and to the right, or down and to the left, as indicated by the two targets marked in the figure. Both possible targets lie well outside the distractor covariance ellipse (which lies at the origin, right on top of the distractor distribution), and we correctly predict efficient search. (B) Search for a stationary target among symmetrically moving distractors. Here the target lies at the center of the covariance ellipse (saliency = 0), and the model correctly predicts a difficult search task.

one direction, so that the two search tasks (search for moving vs. stationary target) were more symmetric. Klempen et al. (1998) have recently demonstrated that in this condition search for a stationary target is quite efficient (RT vs. set size slope of 6 ms/item in target-present trials), as predicted by our model. However, search for a moving target is still more efficient (− 1.2 ms/item). It is quite believable that this lingering asymmetry between search for a moving versus stationary target is due to a basic asymmetry in processing motion. Alternatively, in their experiment observers could see the stationary edges of the monitor, which arguably could act as a stationary distractor and reduce the saliency of a stationary target. What happens in a motion Ganzfeld?

In another infamous asymmetry, either the target moves at a slow speed, while the distractors all move at a faster speed, or vice versa. The target and distractors all oscillate horizontally. Ivry and Cohen (1992) found that search for a fast target was easier than for a slow target. They rule out attributing this effect to 'differences in temporal frequency, discriminability, or one type of representation that might result from spatiotemporal filtering,' and instead suggest that a set of high-pass speed detectors with different low-speed cutoffs would explain the results, since there will be a class of cells that respond to a fast target and not the slow distractors, but no class of cells that will respond to the slow target but not the fast distractors.

We point out that their result makes perfect sense within our framework. Search for a slow target is depicted in Fig. 2A. As in the previous example, the target lies within the covariance ellipse representing the distractor distribution, and therefore search for a slow target should be more difficult than search for a fast

target, in which, as we show in Fig. 2B, the target points lie outside of the covariance ellipse.

In the final classic search asymmetry, Driver et al. (1992) manipulate the heterogeneity of distractor speed and direction to try to determine whether speed and direction are coded independently by the visual system. Following Treisman's (Treisman, 1988) methodology, subjects search for a target defined by a particular attribute, e.g. motion direction, under two conditions. In the Homogeneous condition, the target and distractors all have the same value on some irrelevant attribute, e.g. motion speed. In the heterogeneous condition, the target and distractors take on a range of values in the irrelevant dimension. Treisman suggests that if the relevant attribute is coded independently of the irrelevant attribute, search performance should be the same in the heterogeneous condition as it was in the homogeneous condition. If the two attributes are not coded independently, the variation in the irrelevant attribute should negatively affect performance in the heterogeneous condition.

Driver et al. (1992) use this methodology to test the independence of motion direction and speed in two tasks: one in which observers search for a target of unique speed (speed task), and one in which observers search for a target of unique direction (direction task). In the speed task, results showed significantly slower RTs in the heterogeneous displays, though search was efficient in both cases (4 ms/item in the target-present trials for both conditions). The authors conclude that speed *cannot* be coded independently of direction. In the direction task, search was less efficient than in the speed task in both conditions (14 ms/item in the homogeneous condition, 17 ms/item in the heterogeneous condition), but there was no significant difference be-
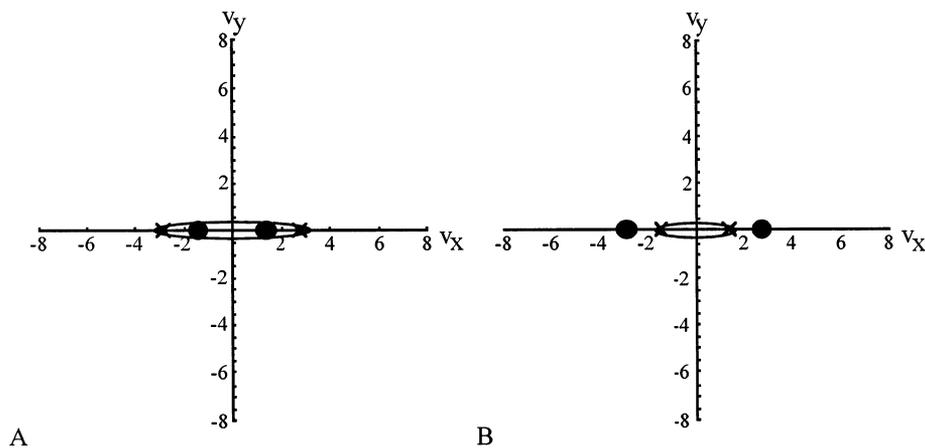
Fig. 2. (A) Search for a slow target (●) among fast distractors (X). Target and distractors oscillate, with random phase. The target lies inside the covariance ellipse, and we correctly predict a difficult search task. (B) Search for a fast target among slow distractors. Target and distractors again oscillate with random phase. Target points lie outside the covariance ellipse, and we correctly predict an easier search task.

tween the target-present data for the two conditions. The authors argue that this could not be attributed to the speeds used being less distinct than the directions, and thus having less of a heterogeneity effect. In the homogeneous conditions for the two tasks, search for a unique speed was more efficient than search for a unique direction, implying that the speeds were *more* discriminable than the directions. The authors conclude that direction *can* be extracted independently of speed.

However, once again our model predicts these results. Viewed in velocity space it becomes clear that the two tasks are not remotely equivalent. In the homogeneous condition of the speed task (Fig. 3A) the saliency (Mahalanobis distance) between the target and mean of the distractors is 4.28. In the heterogeneous condition (Fig. 3B), the saliency was 3.44. Thus, we correctly predict slower search in the heterogeneous condition of the speed task than in the homogeneous condition. One should note that the linear separability model of D'Zmura (1991) and Bauer et al. (1996) would incorrectly predict the same performance in the heterogeneous condition as in the homogeneous condition, since the degree of separability is the same in the two conditions. Also note that both of these saliencies are quite large— we might predict from this that search would be efficient in both cases, as found by Driver et al. (1992).

In the homogeneous condition for the direction task (Fig. 4A), the three Mahalanobis distances, corresponding to the three possible target speeds, are 2.41, 2.53, and 2.57. The mean of these distances, 2.50, serves as a measure of the average ease of search. In the heterogeneous condition for the direction task (Fig. 4B), the Mahalanobis distances for each of the three possible target speeds are 1.31, 2.01, and 3.50, and the mean saliency 2.27. Here we again see a difference between the homogeneous and heterogeneous conditions, but a much smaller difference than in the speed task (2.5–2.3

vs. 4.3–3.4). One could easily believe that there might be no significant difference between the two conditions in the direction task, yet a significant difference between the two conditions in the speed task, as found by Driver et al. (1992).

In addition, recall that Driver et al. find, in the homogeneous conditions, a slope of 4 ms/item for the speed task, and 14 ms/item in the direction task. They take this to indicate that the speeds they used were more discriminable than the directions. This interpretation, in turn, would imply a coarse coding of direction, since the directions used differed by a minimum of 90°. But once again, our simple model in fact predicts this result, since the Mahalanobis distance in the speed task is 4.3, compared with 2.5 in the direction task. Our measure of saliency resembles that used to reject outliers in regression analysis, as discussed in greater detail below. In regression analysis, a point is often rejected as an outlier if this distance is larger than 2.5. This would imply that the speed task would be quite efficient, since its target-distractor distance is much larger than 2.5, while the direction task might fall on the border of inefficient search. Both of these predictions agree with the results of Driver et al.

2.1. Can we also predict color search results?

The question arises whether or not a saliency model can predict search results in other feature dimensions. Color is a particularly interesting example, since like motion it can be thought of as a single, multi-dimensional feature, or split into several 1-D features (e.g. hue, luminance, and saturation), each of which is treated separately, with separate feature detectors. D'Zmura (1991) and Bauer et al. (1996) have suggested the rule that when the target is linearly separable from the distractors, search is easier than when the target
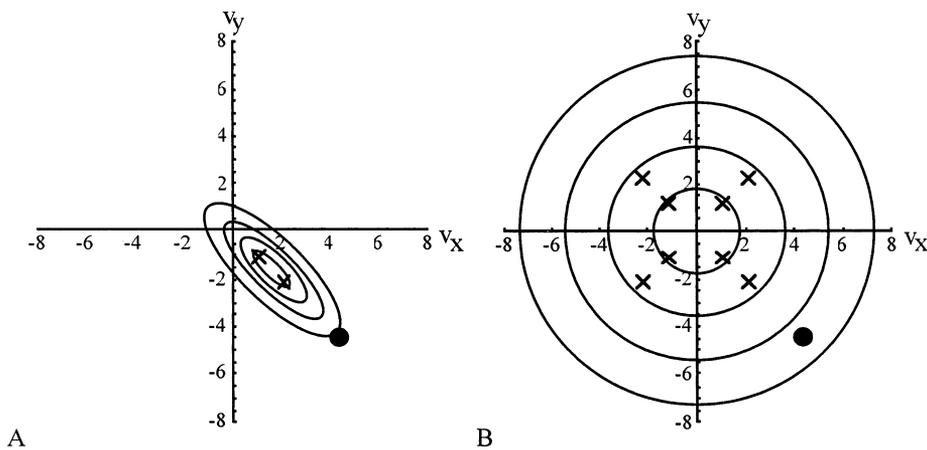
Fig. 3. Speed task. (A) Search for a target of unique speed (●) among distractors of heterogeneous speed, but homogeneous direction (X). The target moved at 6.3 deg/s, with distractors moving at speeds of 3.1 and 1.6 deg/s. Target and distractors moved in the same direction, 45° to either the upper left, upper right, lower left, or lower right. (B) Search for a target of unique speed (●) among distractors of heterogeneous speed and direction. Each distractor motion was randomly chosen from the four directions listed above. The model correctly predicts greater search difficulty in the heterogeneous condition (see text).

and distractors are not linearly separable. Bauer et al. also show there is a continuum, in which search becomes easier the farther the target falls from the separating line. It is trivial to show that our model also predicts their results (in their particular examples, any target not linearly separable from the distractors has zero saliency), and allows us to quantify the improvement in search as the target moves away from the separating line.

For the color experiments of D'Zmura (1991) and Bauer et al. (1996), the linear separability model and the saliency model make essentially the same qualitative predictions. The key difference between the models is that the saliency model cares about the mean and covariance of the distractors, whereas the linear separability model cares only about the existence and location of the line of linear separability, relative to the target. The key for distinguishing between the models is to compare the results of two experiments. In the first experiment, one would search for some target, among some arbitrary collection of distractors. In the second experiment, the experimenter would change the distractors in such a way that the separating line does not move, while the mean and covariance of the distractors changes in such a way as to significantly increase or decrease the saliency of the target. If search results were the same under these two conditions, that would be evidence in favor of the linear separability rule, while if the results changed as predicted by the changed saliency of the target, that would be evidence in favor of the saliency model. Fig. 3 showed an example of such a manipulation, in the motion domain. The results of those motion experiments supported the saliency model and not the linear separability model (see Section 2). Such experiments have yet to be performed in the color domain.

Researchers have also reported color asymmetries much like the motion asymmetries discussed above (Treisman & Gormican, 1988; D'Zmura, 1991; Nagy & Cone, 1996). Nagy and Cone (1996) found that when the target and distractors differed only in saturation, it was easier to detect a saturated target among less-saturated distractors than to detect a less-saturated target among saturated distractors. They found a smaller asymmetry (in the same direction) when target and distractors differed in both saturation and hue, and no asymmetry when target and distractors differed only in hue. (Treisman and Gormican's (1988) found an asymmetry when there was only a hue difference. However, these asymmetries were small, and D'Zmura (1991) and Nagy and Cone (1996) were unable to replicate them.) The saliency model can predict these asymmetries if one considers that the background color may also distract from the target[3]. Both D'Zmura and Nagy and Cone used an unsaturated gray or black background. Fig. 5 depicts the situation. It should be clear that in this situation, search for a saturated target is easier than search for an unsaturated target. Once we include the background as a distractor, the two search tasks are no longer symmetric. We would predict no asymmetry when target and distractors differ only in hue, since inclusion of the background as a distractor does not change the symmetry of the two search tasks.

When the target is not linearly separable from the distractors, but is very different from the distractors, search for that target is very efficient, at least in the case when the observer knows the appearance of the target. This result, which we see in both orientation

---

[3] How many distractors the background should count as, in our model, is an open question.
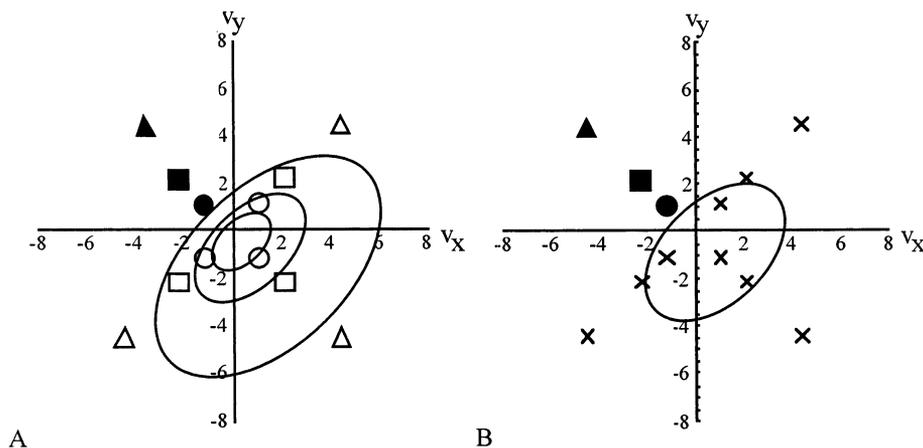
Fig. 4. Direction task. (A) Search for a target of unique direction (filled symbols) among distractors of heterogeneous direction but homogeneous speed (open symbols). The target moved diagonally toward the upper left; the distractors moved toward the lower left, upper right, and lower right. The target was equally likely to be moving at the fast (6.3 deg/s, circles), medium (3.1 deg/s, squares), or slow speeds (1.6 deg/s, triangles). Distractors all moved at the same speed as the target. There are essentially three possible conditions, corresponding to the three possible speeds of the target and distractors. The inner ellipse corresponds to the slowest target and distractors, and the outer ellipse to the fastest. (B) Search for a target of unique direction (filled symbols) among distractors of heterogeneous speed and direction (X). The target moved toward the upper left, and the distractors in the three other possible directions. The target and distractor speeds were each independently chosen from the three speeds listed above. The model correctly predicts very little difference between the homogeneous and heterogeneous conditions for this task (see text).

(Wolfe, Friedman-Hill, Stewart & O'Connell, 1992) and color (Duncan, 1989; Bauer, Jolicoeur & Cowan, 1996) search, cannot be explained by our saliency model for popout, alone. We suggest that, following the popout stage described here, the visual system has a stage which, in the absence of popout in the first stage, acts just like a signal detection stage, as in (Palmer, Ames & Lindsey, 1993). This signal detection stage, for target and distractors that are very different, would easily detect the target element.

3. Discussion

Intuitively, one might expect the visual system to be equipped with a bottom-up mechanism for detecting unusual items in a scene, e.g. items that move in an unusual way when compared with neighboring items. Since such items may be objects that we need to avoid or objects that are otherwise interesting, they should draw our attention automatically, regardless of whether we are explicitly looking for them. For example, if a rock flies at our head, it should draw our attention without our being explicitly on the lookout for rocks flying at our head.

What determines whether an item is unusual? One might expect that the saliency of an item might depend upon the extent to which its feature value (e.g. motion, color, or orientation) was an outlier in the local distribution of feature values, i.e. the extent to which the observation departed from the general pattern of the data set.

We suggest that judging the saliency of an item is equivalent to a parametric test for outliers to a statistical distribution. In a parametric test, a system assumes a given distribution for the feature values, and estimates only the parameters of that distribution. In regression analysis, the distribution is typically assumed to be a normal distribution, and one measure of the degree to which a data point seems to be an outlier is, in 1-D, the number of standard deviations between the point and the mean of the distribution. In higher-dimensional spaces, the degree to which a data point seems to be an outlier is the Mahalanobis distance between the point and the mean of the distribution, i.e. the saliency measure used in this paper. This is not to say that the visual system explicitly assumes that motions in the world are distributed according to a normal distribution. The visual system may use a parametric test for outliers because in a short period of time it is only capable of representing the distribution of motions by their mean and covariance.

We have shown that our model of saliency predicts the results of all of the classic motion search experiments. Previous researchers have explained these same motion popout results on a per-experiment basis: there is an inherent asymmetry in processing of motion information vs. stationary; high-pass speed filters explain the asymmetry in search for a slow versus a fast target; coding of motion direction is coarse; and direction is coded independently of speed but not vice versa. In addition, we have suggested that this model could also predict results in color search experiments.
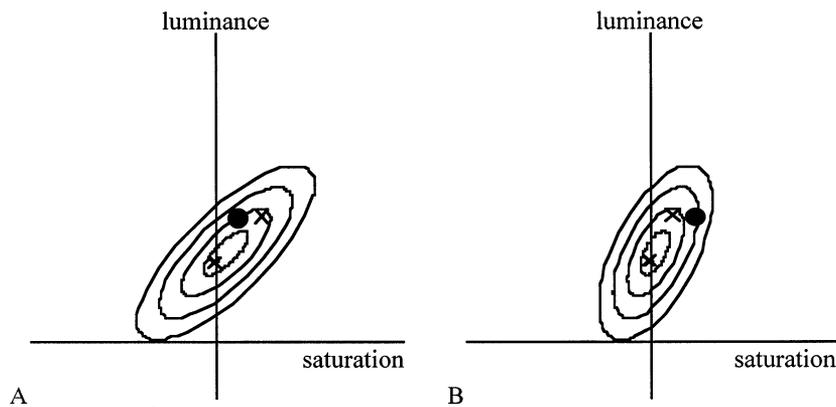
Fig. 5. Sketch of explanation for asymmetry in color search. Saturation is represented by distance from the origin along the *x*-axis, luminance along the *y*-axis. Hue (not shown) is angle in the *x–z* plane. Target and distracting elements (the upper X in both plots) have the same hue, and differ only in saturation. The lower X represents a black background, which also distracts from the target. We weighted the background as if there were four times as many black distractors as others—this arbitrary choice has no qualitative effect on the prediction. (A) Search for a less saturated target among more saturated distractors. (B) Search for a saturated target among less saturated distractors. Note that in (B) the target falls roughly on the 3rd covariance ellipse, whereas in (A) it falls roughly on the 2nd covariance ellipse. Thus we predict that the more saturated target pops out more easily, in agreement with psychophysics.

Researchers often use the results of search experiments to draw conclusions about mechanisms of the visual system. Are there high-pass speed detectors? Is motion direction coded very coarsely? Are motion speed and direction coded independently? Is there a mechanism tuned to orange events (D'Zmura, 1991)? Do we find saturated targets more quickly than unsaturated because saturated signals go through the visual system more quickly (Nagy & Cone, 1996)? However, conclusions one draws about mechanisms from psychophysical experiments are only as good as the underlying model. Our simple model does not directly tell us about visual system mechanisms, but the fact that it predicts the results of a number of search experiments suggests that conclusions about mechanisms, which were drawn from these experiments, need to be reevaluated in light of this model. To draw conclusions from search experiments, one must control for target discriminability. One cannot, for instance, compare a search for a target oriented at 2° among distractors at 0° to search for a red target among blue distractors, and conclude that color search is easy, while orientation search is difficult. At the very least, the success of our model thus far suggests that one control for target saliency, particularly when using the results of search experiments to draw conclusions about visual system mechanisms.

References

Adelson, E. H., & Bergen, J. R. (1985). *Journal of the Optical Society of America*, *2*, 284.

Bauer, B., Jolicoeur, P., & Cowan, W. B. (1996). *Vision Research*, *36*, 1439.

Dick, M. (1989). Thesis, Weizmann Institute.

Dick, M., Ullman, S., & Sagi, D. (1987). *Science*, *237*, 400.

Driver, J., McLeod, P., & Dienes, Z. (1992). *Spatial Vision*, *6*, 133.

Duncan, J. (1989). *Perception*, *18*, 457.

Duncan, J., & Humphreys, G. W. (1989). *Psychological Review*, *96*, 433.

D'Zmura, M. (1991). *Vision Research*, *31*, 951.

Ivry, R., & Cohen, A. (1992). *Journal of Experimental Psychology*: *Human Perception and Performance*, *18*, 1045.

Klempen, N. L., Shulman, E., Royden, C., & Wolfe, J. M. (1998). *Investigative Ophthalmology and Visual Science* (*Supplement*), *39*, 165.

Nagy, A., & Cone, S. M. (1996). *Vision Research*, *36*, 2837.

Nothdurft, H.-C. (1937). *Vision Research*, *33*, 1993.

Palmer, J., Ames, C. T., & Lindsey, D. T. (1993). *Journal of Experimental Psychology*: *Human Perception and Performance*, *19*, 108.

Treisman, A., & Gormican, S. (1988). *Psychological Review*, *95*, 15.

Treisman, A. (1988). *Quarterly Journal of Experimental Psychology*, *40A*, 201.

van Santen, J. P. H., & Sperling, G. (1984). *Journal of the Optical Society of America*, *1*, 451.

Watson, A. B., & Ahumada, A. J. (1985). *Journal of the Optical Society of America A*, *2*, 322.

Wolfe, J. M. (1994). *Psychonomic Bulletin and Review*, *1*, 202.

Wolfe, J. M. (1996). In H. Pashler, *Attention*. London: University College London.

Wolfe, J. M., Friedman-Hill, S. R., Stewart, M. I., & O'Connell, K. M. (1992). *Journal of Experimental Psychology*: *Human Perception and Performance*, *18*, 34.