Current Biology 20, 121–124, January 26, 2010 ©2010 Elsevier Ltd All rights reserved DOI 10.1016/j.cub.2009.11.066

Report

Varying Target Prevalence Reveals Two Dissociable Decision Criteria in Visual Search

Jeremy M. Wolfe[1,2,*] and Michael J. Van Wert[1]
[1]Visual Attention Lab, Brigham and Women's Hospital, 64 Sidney Street, Cambridge, MA 02139, USA
[2]Department of Ophthalmology, Harvard Medical School, 243 Charles Street, Boston, MA 02114, USA

Summary

Target prevalence powerfully influences visual search behavior. In most visual search experiments, targets appear on at least 50% of trials [1–3]. However, when targets are rare (as in medical or airport screening), observers shift response criteria, leading to elevated miss error rates [4, 5]. Observers also speed target-absent responses and may make more motor errors [6]. This could be a speed/accuracy tradeoff with fast, frequent absent responses producing more miss errors. Disproving this hypothesis, our experiment one shows that very high target prevalence (98%) shifts response criteria in the opposite direction, leading to elevated false alarms in a simulated baggage search. However, the very frequent target-present responses are not speeded. Rather, rare target-absent responses are greatly slowed. In experiment two, prevalence was varied sinusoidally over 1000 trials as observers' accuracy and reaction times (RTs) were measured. Observers' criterion and target-absent RTs tracked prevalence. Sensitivity (d′) and target-present RTs did not vary with prevalence [7–9]. These results support a model in which prevalence influences two parameters: a decision criterion governing the series of perceptual decisions about each attended item, and a quitting threshold that governs the timing of target-absent responses. Models in which target prevalence only influences an overall decision criterion are not supported.

Results

Experiment One: High Target Prevalence Elevates False Alarms but Does Not Speed Target-Present Responses

In experiment one, 13 observers performed a simulated baggage search task looking for weapons (guns and knives) that were present on either 50% or 98% of bags. Reaction times (RTs) less than 200 ms or greater than 15,000 ms were excluded. One observer was removed from further analysis for an excess of very fast RTs. For the remaining 12 observers, this led to the removal of 0.5% of trials as outliers.

Figure 1A shows the average error rates for 98% and 50% prevalence. The false-alarm rate increased dramatically from 0.18 at balanced (50%) prevalence to 0.58 at high prevalence in this experiment [$t(11) = 8.0$, $p < 0.0001$]. Miss errors dropped from 0.15 to 0.02 [$t(11) = 8.5$, $p < 0.0001$]. Figure 1B shows the signal detection measures d′ (sensitivity) and c (criterion). d′ was modestly reduced [$t(11) = 2.4$, $p < 0.05$]. However, the use of d′ assumes equal variance of signal and noise distributions. Previous work indicates that this task is better fit by an unequal variance model (as shown in Figure S1C available online, the slope of the z-transformed receiver operating characteristic is about 0.6 rather than the equal variance slope of 1.0 [4]). If corrected for unequal variance, the change in criterion (calculated as Macmillan and Creelman's "C_2"; p. 66 in [10]) remains essentially the same and highly significant.

If the increase in false alarms were the result of a speed/accuracy tradeoff, one might expect target-present RTs to become faster, following the pattern of target-absent RTs at low prevalence. However, as can be seen in Figure 1C, the prevalence manipulation had no effect on either hit [$t(11) = 1.43$, $p = 0.18$] or false-alarm [$t(11) = 0.93$, $p = 0.37$] RTs, disconfirming the tradeoff hypothesis. Interestingly, the only effect on RT that we observed was a massive slowing of target-absent responses [correct absent: $t(11) = 6.67$, $p < 0.0001$; miss: $t(11) = 5.11$, $p < 0.0001$].

Experiment Two: Variable Prevalence Principally Affects Criterion and Target-Absent Reaction Time, Not d′ or Target-Present Reaction Time

In experiment two, 12 observers performed 1000 trials of the simulated baggage search as target prevalence varied sinusoidally from high to low and back to high. RTs less than 200 ms or greater than 15,000 ms were removed as outliers. This removed 0.56% of trials. Trials were binned into 20 blocks of 50 trials each. At very low prevalence, there were very few target-present trials, whereas at very high prevalence, there were very few target-absent trials. We eliminated empty cells from analysis by pooling responses over all 12 observers. For the RT analyses, any cell with fewer than 20 trials across all observers was excluded from analysis.

Figure 2A shows the errors trading off as a function of prevalence. Again, based on evidence that this is an unequal variance task (see Supplemental Experimental Procedures), we calculated d_a as the measure of sensitivity and C_2 as the measure of criterion. Because these statistics are based on pooled data, one should be cautious in interpreting them. Nevertheless, Figure 2B shows that criterion varied systematically with prevalence whereas sensitivity did not. C_2 and prevalence were significantly correlated (Pearson $r = -0.92$; 95% confidence interval [CI]: -0.97 to -0.80, $p < 0.0001$). In contrast, d_a was not systematically related to prevalence (Pearson $r = 0.20$; CI: -0.27 to 0.59, $p = 0.39$). Results do not change markedly if one calculates d′ and c. It is criterion that changes with prevalence. Note that peak criterion value in Figure 2B lagged behind the lowest prevalence. This reflects the number of trials over which the observers based their internal estimates of prevalence. These data do not permit a precise calculation, but it appears that observers compute prevalence over about four dozen trials.

Turning to the RT data, Figure 2C shows that, as in experiment one, it is the target-absent RTs that are clearly responsive to prevalence. Looking at target-present trials (black symbols), it can be seen that both hit and false-alarm RTs decline modestly over the course of experiment. This monotonic trend could represent a general speeding of RT with practice but does not reflect the change in prevalence. The variation in target-absent response times across the

*Correspondence: wolfe@search.bwh.harvard.edu

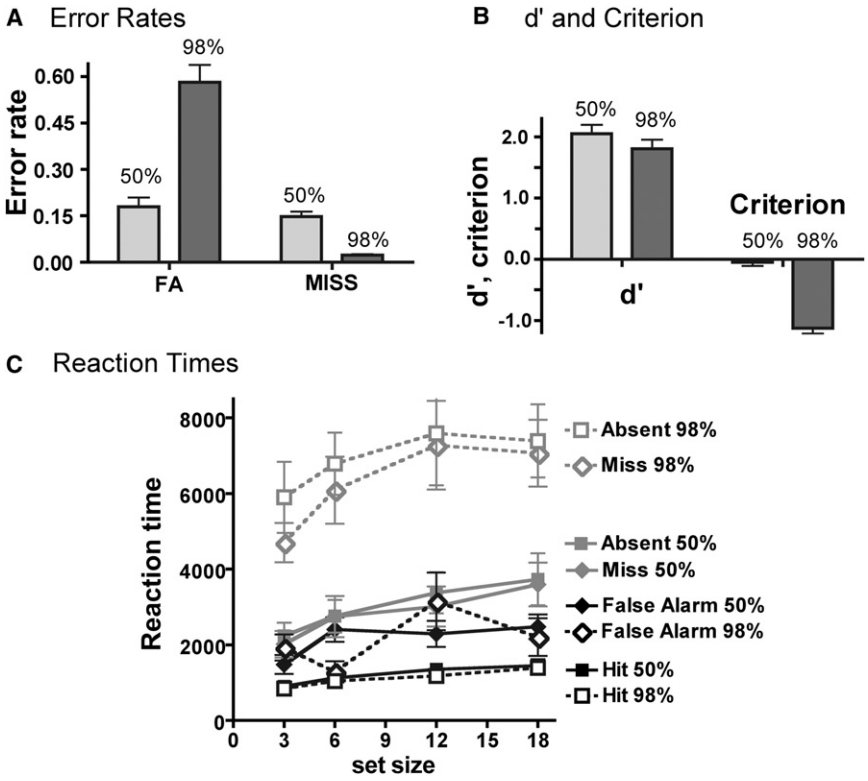

Figure 1. Experiment One: Very High Prevalence Elevates False Alarms and Target-Absent Reaction Times

(A) False-alarm (FA) and miss error rates as a function of target prevalence (50% and 98%).
(B) Signal detection measures: average sensitivity (d′) and criterion (c) values.
(C) Average reaction time (RT) for correct target-present (hit) and target-absent reactions.
Error bars are ± one standard error of the mean.

experiment is about five times greater and more clearly follows prevalence.

Discussion

As anticipated by work in other domains, varying target prevalence causes a tradeoff between false-alarm and miss errors [7–9]. What is novel and informative here is that, for RT, the main effect of prevalence falls on the target-absent responses. Taken together, the pattern of RT and error data falsifies some plausible theories. For example, the pattern of RTs is not consistent with any account holding that RTs are speeded when observers can predict the answer. Were that the case, RTs should be slowest at 50% prevalence and fastest at very high and low prevalence. This is not what the data in Figure 2 show.

A visual search task might be thought of as a two-alternative forced choice (2AFC) decision between a target-present and a target-absent response. 2AFC tasks can be modeled as an accumulation of information toward one of two response boundaries [11–15] (see Figure 3). Errors occur when the noise perturbing the drift toward one boundary causes the accumulation to reach the other boundary by mistake. Our data constrain such diffusion models. Specifically, we argue that modeling the effects of prevalence will require changing more than one parameter. Changing prevalence shifts criterion. To vary criterion in a standard diffusion model, one can move the starting point. In Figure 3, if the starting point moved toward the "yes" boundary at high prevalence, false alarm errors would become more common and misses less common, as desired, without changing sensitivity (represented by the separation between "yes" and "no" boundaries). However, this would also lead to target-present RTs becoming faster and target-absent RTs slower. This speeding of target-present RTs at high prevalence is not seen. A change in the target-absent but not target-present RTs could be produced by moving the "no" boundary. However, because sensitivity (d′ or d_a) varies with the separation between the decision boundaries, moving the "no" boundary down would increase sensitivity at high prevalence, a pattern not seen in our data. (See Supplemental Experimental Procedures for details of simulation of these manipulations of a diffusion model.)

Although the pattern of the data might be captured by simultaneously changing two parameters in a standard diffusion model [12], we adopt a somewhat different approach, the "multiple-decision model," illustrated in Figure 4, because search tasks like ours are not actually simple 2AFC tasks. At any given moment, the observer evaluates some aspect of the display. Figure 4 illustrates the observer selecting a single item. In an "internal decision" stage, the observer makes a 2AFC decision about this information. If the response, R, exceeds a criterion, a target is deemed to be present and the observer makes a "yes" response. If not, the observer continues to search. A second process generates "no," absent responses. This is modeled here as a diffusion toward a quitting threshold. If the diffusion value, Q, exceeds that threshold, a "no" response is generated. Otherwise, a new item is selected and search continues.

Here, the two parameters that are affected by prevalence are the internal decision criterion and the quitting threshold. At high prevalence, criterion moves left, making "yes" responses more likely, and the quitting threshold moves up, making target-absent RTs slower. At low prevalence, the parameters shift in the opposite direction. As shown in Figures S1D and S1E, simulation of a model of this sort produces the basic pattern of results seen in the experiments reported here.

The structure proposed in Figure 4 generalizes quite naturally beyond simple present/absent search tasks and may have some utility in explaining other search phenomena. For example, many radiology tasks require that observers find not one but all targets (e.g., multiple lung nodules). In terms of the model presented here, this means that a "present" response does not end search. The cycle of selection and perceptual decision would continue until the quitting threshold was reached. "Satisfaction of search" is a known problem in search for an unknown number of targets [16, 17]. This is the observation that the probability of detecting one target is lower if another target has been detected first. This phenomenon could be a consequence of the dual-threshold nature of search. Suppose that two trials have the same quitting

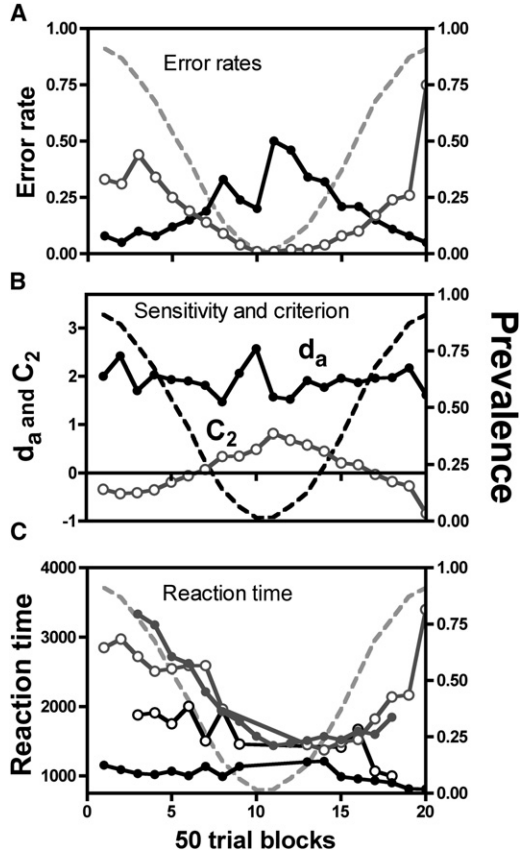

Figure 2. Experiment Two: Changing Target Prevalence Changes the Pattern of Errors and Target-Absent Reaction Times

(A) Miss errors (solid black symbols) and false-alarm errors (open gray symbols) trade off as prevalence (dashed line) varies over 1000 trials.

(B) d_a (solid black symbols), a signal detection measure of sensitivity, does not vary systematically with prevalence, but C_2 (open gray symbols), a criterion measure, does.

(C) Hit RTs (solid black symbols) change very little with prevalence, whereas true negative responses (open gray symbols) vary markedly. False-alarm errors (open black symbols) do not vary with prevalence, though they appear to become faster during the experiment. Miss errors (solid gray symbols) vary with prevalence in a manner similar to true negatives. (See also Figures S1A–S1C.)

threshold. In one trial, the image contains T1 and T2. In the other, only T2 is present. If we suppose that it takes some time to deal with T1 and that the quitting threshold discounts this fact, then the chance of reaching T2 will be lower on the T1 + T2 trial than on the T2 alone trial. Further research would be needed to test this hypothetical account of satisfaction of search, but the account does capture the possibility of a separation between finding a target and ending a search.

Experimental Procedures

Experiment One

Participants

Thirteen paid participants between the ages of 18 and 55 were tested in all conditions. Each participant reported no history of eye or muscle disorders. All had 20/25 or better vision and passed Ishihara's tests for color blindness. Informed consent was obtained for all participants, and each participant was paid US $10/hour.

Stimuli

Realistic bag stimuli were created by placing X-ray images of assorted objects in X-ray images of empty bags. Items were semitransparent and could overlap. Component bags and objects were X-ray images provided by the Transportation Security Laboratory of the United States Department of Homeland Security. Set size was varied by varying the number of items added to the bag (3, 6, 12, or 18). Bags and individual objects were scaled in an appropriate manner so, as an example, a computer would be bigger than an iPod. Observers sat at approximately 57 cm from the screen. At this distance, bags subtended a range of sizes 9.5° in height × 16° in width to 20° in height × 21.5° in width. Eight pieces of clothing were added to each bag but were not counted in the set size. In these images, clothing adds an indistinctly shaped orange haze to the image. Stimuli were presented on

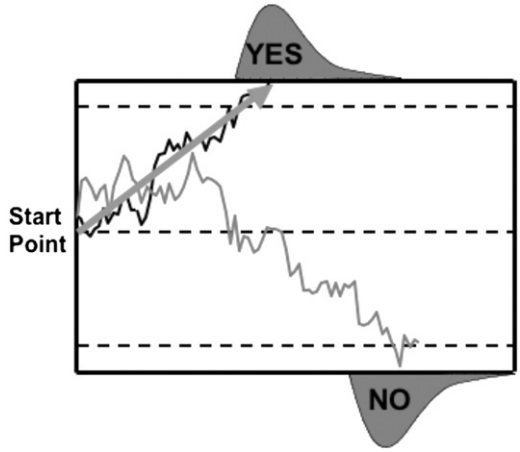

Figure 3. The Drift Diffusion Model

In a standard drift diffusion account of a two-alternative forced choice (2AFC) task, information begins accumulating at a start point. It generates one response (here, "yes") if it reaches an upper bound and another ("no") if it reaches a lower bound. For a fixed drift rate, sensitivity (d′) can be varied by varying the separation of the bounds, and criterion can be varied by changing the starting point. (See also Figures S1D and S1E.)

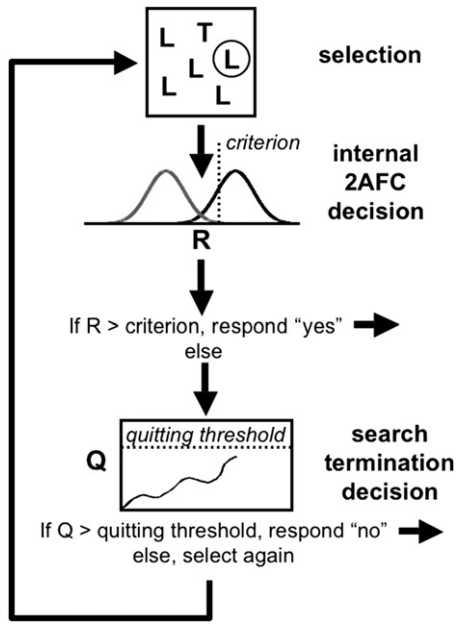

Figure 4. A Multiple-Decision Model for Visual Search

In this model, the observer makes a 2AFC decision about each item that is selected. If an item is classified as a target, a "yes" response is generated. If not, a new item will be selected unless a target-absent decision is generated when a quitting signal exceeds its threshold. The quitting signal is modeled as a diffusion process. (See also Figures S1D and S1E.)

Macintosh computers running MATLAB 7.5 with the Psychophysics Toolbox, version 3 (http://psychtoolbox.org/; [18, 19]).

Procedure

To familiarize observers with the threat stimuli, they were first briefly shown 20 examples of weapons for 1 s in isolation. Next, they were given 100 practice trials at 50% prevalence with full feedback on the correctness of responses. Observers were instructed to indicate as quickly and accurately as possible whether a target was present or absent. On each trial, a fixation cross and audible click were followed after 200 ms by the stimulus. The stimulus remained visible until the observer responded. A 500 ms blank interval preceded the start of the next trial.

After practice, observers completed the two experimental blocks: 200 trials at 50% prevalence and 1000 trials at 98% prevalence. Order of the two blocks was counterbalanced over observers. Observers were told that bags without weapons would be "frequent" in the 50% prevalence condition, and that bags without weapons would be "rare" in the 98% prevalence condition. We emphasized that they should try to be as quick and accurate as possible in correctly identifying bags without weapons. Full feedback was given after each trial. If a target was present, it was outlined with a box and shown to the observer. A 2 min break was enforced every 200 trials (about every 20 min).

Experiment Two

Participants

Twelve paid participants between the ages of 18 and 55 were tested in all conditions. Each participant reported no history of eye or muscle disorders. All had 20/25 or better vision and passed Ishihara's tests for color blindness. Informed consent was obtained for all participants, and each participant was paid US $10/hour.

Procedure

The stimuli and general methods were essentially identical to those of experiment one. Observers were familiarized with the targets in advance and were then tested for 100 trials of training at 50% prevalence with full feedback. Finally, over the course of a block of 1000 trials with full feedback, prevalence varied sinusoidally through one cycle from 100% on trial 1 to 0% at trial 500 and back to 100% by trial 1000. Any given trial could be target present or target absent, with the probability of target presence determined by the current prevalence. Observers were told that the probability of a target would vary over time. A 2 min break was enforced after every 200 trials (about every 20 min).

Supplemental Information

Supplemental Information includes one figure and Supplemental Experimental Procedures and can be found with this article online at doi:10.1016/j.cub.2009.11.066.

Acknowledgments

This research was supported by grants to J.M.W. from the National Institutes of Health and National Eye Institute (EY017001) and from the United States Department of Homeland Security (02-G-010). We thank P. Howe and T. Horowitz for advice.

Received: May 24, 2009
Revised: November 4, 2009
Accepted: November 5, 2009
Published online: January 14, 2010

References

1. Wolfe, J.M. (1998). Visual search. In Attention, H. Pashler, ed. (Hove, UK: Psychology Press), pp. 13–74.
2. Verghese, P. (2001). Visual search and attention: A signal detection theory approach. Neuron *31*, 523–535.
3. Wolfe, J.M., and Reynolds, J.H. (2008). Visual search. In The Senses: A Comprehensive Reference, *Volume 2*, A.I. Basbaum, A. Kaneko, G.M. Shepherd, and G. Westheimer, eds. (San Diego, CA: Academic Press), pp. 275–280.
4. Wolfe, J.M., Horowitz, T.S., Van Wert, M.J., Kenner, N.M., Place, S.S., and Kibbi, N. (2007). Low target prevalence is a stubborn source of errors in visual search tasks. J. Exp. Psychol. Gen. *136*, 623–638.
5. Wolfe, J.M., Horowitz, T.S., and Kenner, N.M. (2005). Cognitive psychology: Rare items often missed in visual searches. Nature *435*, 439–440.
6. Fleck, M.S., and Mitroff, S.R. (2007). Rare targets are rarely missed in correctable search. Psychol. Sci. *18*, 943–947.
7. Healy, A.F., and Kubovy, M. (1981). Probability matching and the formation of conservative decision rules in a numerical analog of signal detection. J. Exp. Psychol. Hum. Learn. *7*, 344–354.
8. Treisman, M. (1984). A theory of criterion setting: An alternative to the attention band and response ratio hypotheses in magnitude estimation and cross-modality matching. J. Exp. Psychol. Gen. *113*, 443–463.
9. Maddox, W.T. (2002). Toward a unified theory of decision criterion learning in perceptual categorization. J. Exp. Anal. Behav. *78*, 567–595.
10. Macmillan, N.A., and Creelman, C.D. (2005). Detection Theory (Mahwah, NJ: Lawrence Erlbaum Associates).
11. Reeves, A., Santhi, N., and Decaro, S. (2005). A random-ray model for speed and accuracy in perceptual experiments. Spat. Vis. *18*, 73–83.
12. Palmer, J., Huk, A.C., and Shadlen, M.N. (2005). The effect of stimulus strength on the speed and accuracy of a perceptual decision. J. Vis. *5*, 376–404.
13. Brown, S.D., and Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. Cognit. Psychol. *57*, 153–178.
14. Ratcliff, R. (2006). Modeling response signal and response time data. Cognit. Psychol. *53*, 195–237.
15. Ratcliff, R. (1978). A theory of memory retrieval. Psychol. Rev. *85*, 59–108.
16. Berbaum, K.S., Franken, E.A., Jr., Dorfman, D.D., Rooholamini, S.A., Kathol, M.H., Barloon, T.J., Behlke, F.M., Sato, Y., Lu, C.H., el-Khoury, G.Y., et al. (1990). Satisfaction of search in diagnostic radiology. Invest. Radiol. *25*, 133–140.
17. Nodine, C.F., Krupinski, E.A., Kundel, H.L., Toto, L., and Herman, G.T. (1992). Satisfaction of search (SOS). Invest. Radiol. *27*, 571–573.
18. Brainard, D.H. (1997). The Psychophysics Toolbox. Spat. Vis. *10*, 433–436.
19. Pelli, D.G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. Spat. Vis. *10*, 437–442.