Optimal feature integration in visual search

Benjamin T. Vincent School of Psychology, University of Dundee, Dundee, UK

Roland J. Baddeley Department of Experimental Psychology, University of Bristol, Bristol, UK

Tom Troscianko Department of Experimental Psychology, University of Bristol, Bristol, UK

Iain D. Gilchrist Department of Experimental Psychology, University of Bristol, Bristol, UK

Despite embodying fundamentally different assumptions about attentional allocation, a wide range of popular models of attention include a max-of-outputs mechanism for selection. Within these models, attention is directed to the items with the most extreme-value along a perceptual dimension via, for example, a winner-take-all mechanism. From the detection theoretic approach, this MAX-observer can be optimal under specific situations, however in distracter heterogeneity manipulations or in natural visual scenes this is not always the case. We derive a Bayesian maximum a posteriori (MAP)-observer, which is optimal in both these situations. While it retains a form of the max-of-outputs mechanism, it is based on the maximum a posterior probability dimension, instead of a perceptual dimension. To test this model we investigated human visual search performance using a yes/no procedure while adding external orientation uncertainty to distracter elements. The results are much better fitted by the predictions of a MAP observer than a MAX observer. We conclude a max-like mechanism may well underlie the allocation of visual attention, but this is based upon a probability dimension, not a perceptual dimension.

Keywords: 2-AFC, yes/no, detection, visual search, attention, Bayes, signal detection theory, target visibility, optimal observer, Monte Carlo, internal noise, external noise

Citation: Vincent, B. T., Baddeley, R. J., Troscianko, T., & Gilchrist, I. D. (2009). Optimal feature integration in visual search. *Journal of Vision, 9*(5):15, 1–11, http://journalofvision.org/9/5/15/, doi:10.1167/9.5.15.

Introduction

While it is undoubtedly the case that high-level factors such as scene context (Henderson, 2003; Torralba, 2003) and task demands (Buswell, 1935; Land & Hayhoe, 2001; Yarbus, 1967) play an important role in natural visually guided behaviors, it is also well known that simple visual features can affect the allocation of visual attention. For example in visual search, search efficiency is dependent upon the feature dimensions used (Wolfe, 1998) and the heterogeneity of the distracters (Duncan & Humphreys, 1989, 1992; Pashler, 1987). These phenomena require an explanation: in what way are basic visual cues utilized to guide visual attention?

There has been an abundance of approaches and models applied to this problem. While not universally accepted (Allport, 1989), many approaches such as Feature Integration Theory (Treisman & Gelade, 1980) and Guided Search (Wolfe, 1994), are strongly influenced by the notions of finite processing capacity, internal selection processes, and serial deployment of attention. A key experimental phenomenon supporting this notion was the set size effect: for features not classed as 'basic' (Wolfe & Horowitz, 2004) increasing the number of display items took longer to search (the set size effect). Despite various advantages, studies that use long presentation times and primarily use reaction time as a dependent variable have the disadvantage that eye movements can be made during the search. Unless eye trackers are used, one could not control for, nor eliminate, their influence on reaction times. For example, it is known for many features, including oriented lines, that detectability dramatically declines with retinal eccentricity (Carrasco & Frieder, 1997) and so Zelinsky and Sheinberg (1997) argued that inferences about internal search processes are confounded with eye movements unless these are specifically accounted for. Controlling for retinal location of stimuli, combined with the notion of detectability (e.g. Verghese & Nakayama, 1993) underlies an alternative approach.

The detection theoretic approach was applied to visual search by Palmer, Ames, and Lindsey (1993) whose mathematical terminology we partially use here. It uses short stimulus presentation times so as to control for eye

doi: 10.1167/9.5.15 Received October 7, 2008; published May 15, 2009 ISSN 1534-7362

movements and records performance measures as the dependent variable, see review by Verghese (2001). This approach has 3 main steps: 1) representation of the stimuli; 2) a combination rule which utilizes the visual information; 3) a decision rule that determines behavioral response.

This signal detection theory account which we describe below has provided close matches to human visual search performance, shedding light on a variety of empirical phenomena. These include: set size effects (Cameron, Tai, Eckstein, & Carrasco, 2004; Eckstein, 1998; Eckstein, Thomas, Palmer, & Shimozaki, 2000; Palmer, 1994; Palmer et al., 1993) and conjunction searches where the target is defined by a combination of 'basic' features present in the distracters (Eckstein, 1998; Eckstein et al., 2000). Predictions for target distracter similarity manipulations, effects of multiple targets and external noise have also been calculated (Palmer, Verghese, & Pavel, 2000). While the detection theoretic approach applied to visual search thus far stops at this decision stage, it is still a model of attentional allocation, and there is no a priori reason why it cannot be extended to longer display duration. Search performance in such detection tasks has been related to search time (Palmer, 1998), but further work in this area is needed (see Najemnik & Geisler, 2005).

Many detection theoretic applications to visual search have examined performance on a 2-TAFC (Temporal Alternative Forced Choice) task where subjects respond which of 2 successive stimulus displays contains a target. Many of these also provide at least appendices on how the approach applies to yes/no detection: responding if a target is present or absent in a single stimulus presentation. While the TAFC task has the advantage of only requiring calculation of a percent correct measure, it has the potential to inadvertently involve memory mechanisms if the delay between successive stimulus presentations is long. For example, Wilken and Ma (2004) apply SDT to a 2-TAFC design with the specific aim of studying visual short term memory. Here, we avoid this possibility by examining visual search performance with the yes/no task.

Below, we describe the signal detection theory account and show that while it has successfully accounted for many phenomena, its assumptions limit its robustness, sometimes resulting in suboptimal predicted performance levels. Therefore, we develop a Bayes-optimal observer which maintains optimal performance in situations where the detection theoretic model does not. The major difference is that the Bayes-optimal observer makes decisions based upon maximum a posteriori *probability* (hence the name MAP-observer) as opposed to the maximum observation along a *perceptual* dimension. Our results support the notion that we are near Bayesian-optimal observers, making decisions based on a probability axis, rather than a perceptual one.

Signal detection theory for yes/no task (the MAX observer)

Step 1—Representation

While target and distracter items take on specific experimentally defined values (v) such as orientation, our internal estimate (i.e. percept) of these (u) stimuli values has a degree of uncertainty. The uncertainty will take on specific extents, thus our representations of targets (T) and distracters (D) can be described by Gaussian distributions[1] $u_t = N(\mu_T, \sigma_T)$ and $u_d = N(\mu_D, \sigma_D)$ (Green & Swets, 1966). In this way our representations accurately capture both the average percept associated with a stimulus and also its degree of variability.

Step 2—Integration

In standard (and most commonly used) detection theory, a MAX combination rule is used. In the yes/no detection task, 1 presentation is given that can either be a signal (s) or noise trial (n). The internal percepts of the N display items in the target present trial can be described as $U_s = \{t_1, d_1, \ldots d_{N-1}\}$ while the target absent trial as vector $U_n = \{d_1, \ldots d_N\}$. According to the MAX rule, only the highest valued percept is considered (i.e. $\max(U_n)$ and $\max(U_s)$) and then passed on for the last step of decision making.

Step 3—Decision

A decision of whether the target was present or absent in a given trial is determined by comparing the maximum of internal responses on the trial to an internal threshold criterion, c. The probability of deciding the target is present in a noise trial (i.e. false alarm rate) is simply the proportion of maximum percept above a criterion on noise trials $P(\text{'yes'}\,|\,n) = P(\max(U_n) > c)$. Similarly the probability of deciding the target is present in a signal trial (i.e. hit rate) is the proportion of maximum percepts above a criterion on signal trials $P(\text{'yes'}\,|\,s) = P(\max(U_s) > c)$, see Palmer et al. (2000). By moving the models' internal criterion c, it is possible to trace out a predicted ROC curve: this can be compared to empirically collected hit and false alarm rates in yes/no visual search experiments.

This account is suboptimal when distracters are heterogeneous

The detection theoretic approach to visual search described above (i.e. a MAX observer) will make optimal

decisions about target presence and absence: but only under specific conditions. In particular, when the degree of uncertainty of noisy observations of targets and distracters are identical (and Gaussian distributed) then the MAX observer will perform as well as possible. The majority of detection theoretic studies of visual search studies have used searches which do conform (or closely approximate) this particular situation (see Figure 1a). An intuitive way to understand why the MAX observer is optimal in this situation, is to see that given an observation, the probability that it was a target monotonically increases with the value of the observation (Figure 1a; see Appendix A for more details).

One situation which will most obviously violate the narrow conditions of optimality of the MAX observer is an experiment in which distracter heterogeneity is manipulated. Here, distracters are often pseudo-randomly chosen from a number of distinct feature values (e.g. 40°, 45°, 50° oriented lines) such that the observer has increased uncertainty about what features are observed in the presence of distracters. Alternatively one can, as we do in this study, inject continuous external Gaussian distributed noise onto the feature property (i.e. orientation) displayed on the screen associated with distracters (see Figure 1c). Now, the MAX observer will no longer optimally decide on presence or absence because given an observation, the probability it was caused by a target item no longer monotonically increases with increasing values of the sensory observation (Figure 1d).

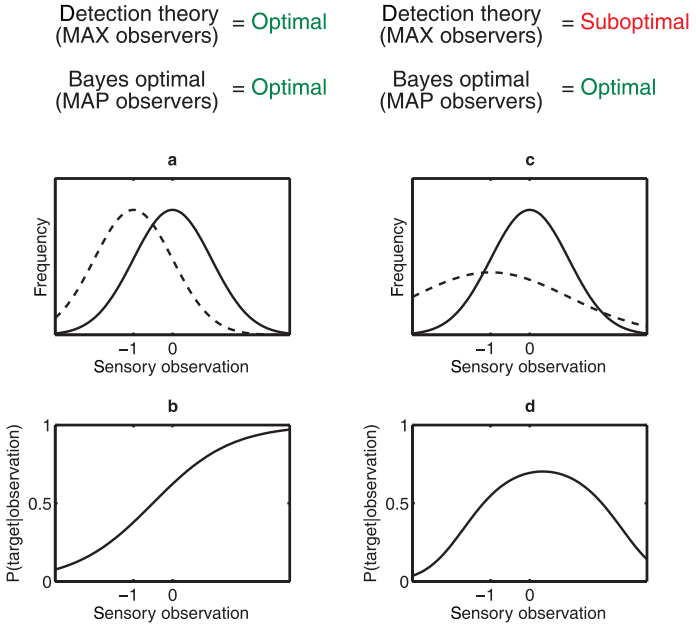

Figure 1. In the case where uncertainty about an observed feature value of both target and distracter items is equal (left column) then the MAX and MAP predictions are identical (left column; a). In this case, the probability of observing a target monotonically increases with the sensory observed value increases (b). When target and distracter uncertainty is unequal (right column; c), MAX observers result in suboptimal predictions because target probability is no longer monotonically related to an increase in sensory observation (d). The MAP observer is optimal in both situations.

Rosenholtz (2001) compared human performance to predictions of SDT and 2 close variants, for visual searches for oriented lines. Targets always took on a specific and unchanging orientation. Distracters however, could take on either 1, 2 or 3 specific orientations (depending on the experiment). Within an experiment, distracters could take on particular orientations with a set of probabilities (depending on condition). Overall, correlations between human and predicted performance were fair, in the range r = 0.68 to r = 0.87, but left plenty of scope for improvements to be made to the model. Rosenholtz suggested this may be done through accounting for difficulty of representing targets.

An additional reason why model fits left room for improvement is that the signal detection theory (i.e. MAX observer) is suboptimal in the case where there is unequal uncertainty in target and distracter properties. If humans are optimal observers, clearly the MAX observers good fit to human data in previous studies will not extend to manipulations of distracter heterogeneity. In this study we use a distracter heterogeneity manipulation and show that human performance is better described as being near Bayesian-optimal than by the MAX observer described above. Before we can show this, we derive the optimal Bayesian observer (a MAP observer). The essential difference is that the MAP observer makes decisions based upon the probability that any observation is to actually be a target, as opposed to the MAX observer which makes decisions upon a perceptual feature dimension.

Bayesian optimal observer for yes/no task (the MAP observer)

Because the yes/no task involves two mutually exclusive and exhaustive hypotheses (target present or absent), it has been customary to formulate optimal observers in terms of a likelihood ratio (Green & Swets, 1966; Wickens, 2002). This has certain advantages and we detail it in Appendix A, but for our purposes we use Bayes equation and detail the account below. This makes no difference to the predictions, and the way of describing the model only has an impact when considering neural implementation of such an account (see later). Ideal observers for the yes/no task have been used previously; see Palmer et al. (2000) and Shimozaki, Eckstein, and Abbey (2003).

Step 1: Representation of targets and distracters

Targets and distracters are represented in the same way as before. However because we later invoke Bayes equation these representations are the likelihoods $P(u|T)$ and $P(u|D)$, that is the probability of an internal percept given the display item is a target or distracter, respectively. For our purposes, it is also useful to define two separate sources of uncertainty: one external source of real uncertainty in the display and distracter items, σ_T and σ_D. The second source is internal noise σ_N, produced for example by neural noise (e.g. Tolhurst, Movshon, & Dean, 1983). We make the simplifying assumption that internal noise is homogeneous over certain ranges of stimulus values (v). Therefore, resulting target and distracter distributions respectively are $P(u|T) = \mathrm{N}(\mu_T, \sqrt{\sigma_T^2 + \sigma_N^2})$ and $P(u|D) = \mathrm{N}(\mu_D, \sqrt{\sigma_D^2 + \sigma_N^2})$. In words, target and distracter representations not only capture the actual variability in the orientation displayed in an experiment, but also the degree of internal noise of estimating the actual orientation.

Step 2: Bayesian optimal combination rule

Rather than calculating the max of sensory percepts, the posterior probability of each display item is calculated, and the maximum of these values is taken (i.e. maximum a posteriori, MAP). MAP_s defines a vector of posterior probabilities observed on signal trials $MAP_s = \max(P(T|U_s))$ and similarly MAP_n for noise trials $MAP_n = \max(P(T|U_n))$. While this may seem a trivial difference, it is this which determines if feature information is utilized in an optimal way or not. The question then is how to calculate these posterior probabilities $P(T|U)$. This can be done simply by using Bayes equation

$$P(T|u) = \frac{P(u|T)}{P(u)} P(T), \quad (1)$$

where the evidence term $P(u)$ describes a 'weighted sum' of target and distracter likelihoods $P(u) = P(u|T) \cdot P(T) + P(u|D) \cdot P(D)$. This can be thought of as a distribution representing the relative occurrence of features observed.

The Bayesian prior probabilities $P(T)$ and $P(D)$ define the relative occurrence of targets and distracters. On target present trials, only $1/N$ of the display items is the target. However the prior probability of the target appearing is not $1/N$ because target present trials only occur half of the time. Therefore the prior probability of a display item being a target is $P(T) = 1/2N$. And correspondingly, the rest of the display items will be distracters $P(D) = 1 - P(T)$. For example, if the set size $N = 4$ then $P(T) = 1/8$ and $P(D) = 7/8$.

Step 3: Decision

The decision rule is exactly the same as for the SDT account described above, except now we compare the MAP values against the criterion rather than the max of the internal percepts. False alarm rates can be determined by $P(\text{'yes'}|n) = P(MAP_n > c)$ and hit rates by $P(\text{'yes'}|s) = P(MAP_s > c)$ and by varying the criterion c we obtain an ROC curve.

Monte Carlo simulated trials

Previous investigators have calculated predicted performance for the max rule in the AFC task and the yes/no task by developing analytical expressions. The advantages are that the evaluation is tractable because of the simple Gaussian distribution of target and distracter percepts, and they result in precise predicted performance levels. However these may be troublesome to calculate if dealing with arbitrary target and distracter feature distributions. In our calculation of predicted performance for MAX and MAP observers, we use the process of Monte Carlo simulated trials which is summarized in Figure 2. This has the advantage of more intuitively representing the modus operandi of observers during real trials. Calculation of performance asymptotic with analytically derived results can be achieved by using a large numbers of simulated trials: we used 500,000 simulated trials per data point. As

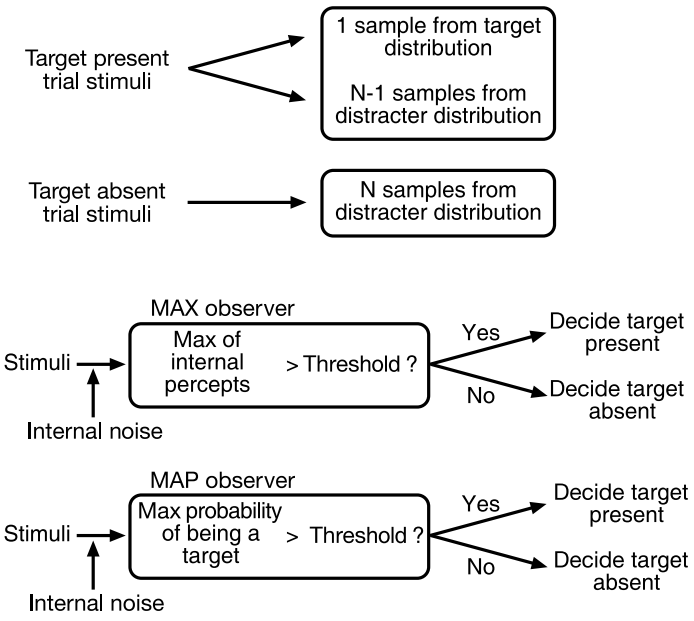

Figure 2. Summary of generating simulated Monte Carlo visual search stimuli for a set size of *N,* and how MAX and MAP observers make decisions on each simulated trial. By repeating this process many times, one can determine hit and false alarm rate which then result in ROC curves.

can be seen in Figure 2, the major difference is that the MAP observer makes decisions based upon a probability dimension rather than a perceptual dimension.

Application to our task

In many search tasks for orientation, the mean target and distracter orientations are different, and the magnitude of this difference is intuitively related to task difficulty. In the particular search we employed, distracters had an orientation of 0° (vertical) and distracters had a *mean* orientation of 0°. In the case of no added distracter orientation uncertainty (Figures 3a and 3b), this task cannot be performed above chance (i.e. $d' = 0$). However, as an increasing amount of distracter uncertainty is applied (Figures 3c, 3d, 3e, and 3f) then it becomes more probable that the distracter orientations are different from target orientations. This prediction of increasing performance with increasing distracter heterogeneity is the opposite trend to that seen in previous reports using the reaction time paradigm (e.g. Duncan & Humphreys, 1989, 1992); however this is specifically due to the fact that in our task the target and distracter orientations have identical mean values. Pilot (unpublished) data does in fact show that if target and distracters have different means, performance does decrease with increasing distracter heterogeneity, in line with previous reports. The advantage of our choice of identical target and distracter means combined with high distracter variability is that distracter orientation can be both higher and lower than the target orientation, so is ideally suited in distinguishing MAX and MAP observer hypotheses.

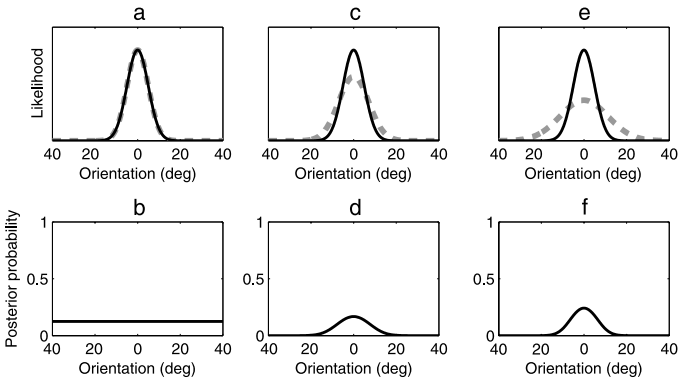

Figure 3. The MAP observer for our task where the mean distracter is identical to the target orientation (N = 4; $\mu_T = \mu_D = 0°$). When no external distracter orientation noise is applied (a), the uncertainty of targets (solid black line) and distracters (dashed gray line) are equal ($\sigma_T = \sigma_D = 0°$, $\sigma_N = 5°$). This task cannot be performed above chance (b). As distracter heterogeneity is increased (c, $\sigma_D = 10°$ to e, $\sigma_D = 20°$) targets and distracters become more distinguishable, and the informativeness of observing a particular orientation upon target/distracter identity changes (d, f).

Methods

Participants

Three subjects with normal or corrected to normal vision participated in this study without payment. Subjects were instructed that response time was unimportant. They were also informed that by chance target presence or absence may repeat a number of times, but that they should base their response solely upon the visual features present on a trial.

Apparatus

Images were displayed on a CRT monitor with resolution set to 1024 × 768 and refresh rate of 85 Hz. The room was darkened to avoid reflections on the monitor or unequal luminance levels due to external sunlight level changes. Viewers used a chin-rest resulting in a stable viewing distance of 52 cm. The screen subtended 38.2° × 29.2°, with ~26 pixels/°.

Stimuli

All stimuli were Gabors with spatial frequency of 4 cpd, contrast of the full range of the monitor and the Gaussian envelope had a standard deviation of 0.22°. Targets were always oriented vertically with no experimentally induced variation ($\mu_T = 0°$, $\sigma_T = 0°$). Distracter orientation was sampled from a Gaussian distribution with a mean of 0° and standard deviation σ_D which was experimentally manipulated. This reflects the degree of externally added orientation noise. To clarify, this results in a heterogeneous set of distracters on each individual stimulus presentation. All display items were presented at constant retinal eccentricity of 11.2° from screen center to minimize eccentricity based variation in detection probability. A set size of N = 4 display items was used, items were presented on the diagonals. Use of 4 display items avoids any potential crowding or lateral interaction effects as inter-item distance is high. This is important as it would violate the assumption that each display items in an independent observation. On target absent trials 4 distracters were present, on present trials, 1 target and 3 distracters were present.

Procedure

A yes/no detection procedure was used. A small fixation blob was continuously present at screen center, avoiding problems maintaining accurate vergence on an otherwise homogenous screen lacking in depth cues. A pre-trial interval of 1000 ms was used. On each trial the target was either present or absent, pre-computed to ensure 50% of each, presented in randomized order. The stimuli for each

trial were presented for 94 ms (a multiple of the screen refresh interval) before being removed, thus controlling for eye movements during stimulus display. One disadvantage of the yes/no task is that an internal criterion value needs to be systematically manipulated in order to trace out an ROC curve. This requires many trials and highly practiced observers. In preliminary experiments we found the use of a rating procedure (Wickens, 2002, chapter 5) avoided these problems and provided accurate and repeatable ROC curves. In this procedure, detailed in Wickens (2002; chapter 5), if subjects thought the target was absent they responded with a key-press of "Q" confident, "W" reasonably confident, "E", not confident and if they thought it was present they responded with "P", confident, "O" reasonably confident, "I", not confident. We recorded the frequency of each response in signal and noise trials. A session consisted of 7 blocks of 40 trials, resulting in 280 trials for each condition. The experiment started with a practice block. Subject BV had many hours of practice sessions over many days, CS had a 1 hour practice session on a day prior to testing, GH had no practice sessions.

Analysis

In every trial, the participant responds with 1 of 6 responses (see above), which can be modeled as a multinomial process. For each condition, this resulted in 5 points along an ROC curve (Wickens, 2002) and the trapezoidal area under the ROC curve (AUC) was calculated. This amounts to the maximum likelihood estimate of the AUC. In order to calculate confidence intervals in the AUC measure, 1000 samples from a Dirichlet distribution (the conjugate prior of the multinomial—its parameters were the rating frequencies) were taken, then converted to FAR and HR values and AUC values of all 1000 samples were calculated. We report mean and 95% confidence of these values (Figure 5). This approach takes the non-independent errors of each FAR/HR point into account.

In order to determine the best fitting internal noise parameter we used leave-one-out cross validation (based on experimental conditions) and pick the internal noise value with lowest mean training set error, but report mean test-set error. Fits are RMS errors between predicted and human AUC values, averaged over each cross validation fold.

Results

The generalization performance (based on leave-one-out cross validation) of the MAP and MAX observers to account for human performance was calculated. This goodness of fit was calculated over many values of the single free parameter of both models, the internal noise (Figure 4, left column). The MAX observer provides a consistently poor fit, whereas the MAP observer provides a good fit to the data over a wide range of parameter values. The possible exception to this is in the case of nearly no internal noise (for subject BV) where the MAP observer predicts abnormally high levels of performance (not shown).

Furthermore, for each observer, there is a clear range of internal noise values which fits the human performance data over all conditions very well (best fitting σ_N values based on test set performance: BV 5.9°, CS 4.4°, GH 4.9°). From the goodness of fit (Figure 4, left column), we can conclude that the MAP observer provides a much better account than the MAX observer for all realistic situations (i.e. when internal noise is not entirely absent). But how well do the predictions fit performance as a function of distracter heterogeneity?

Using the value of the internal noise parameter corresponding to best training set fit (which is of the

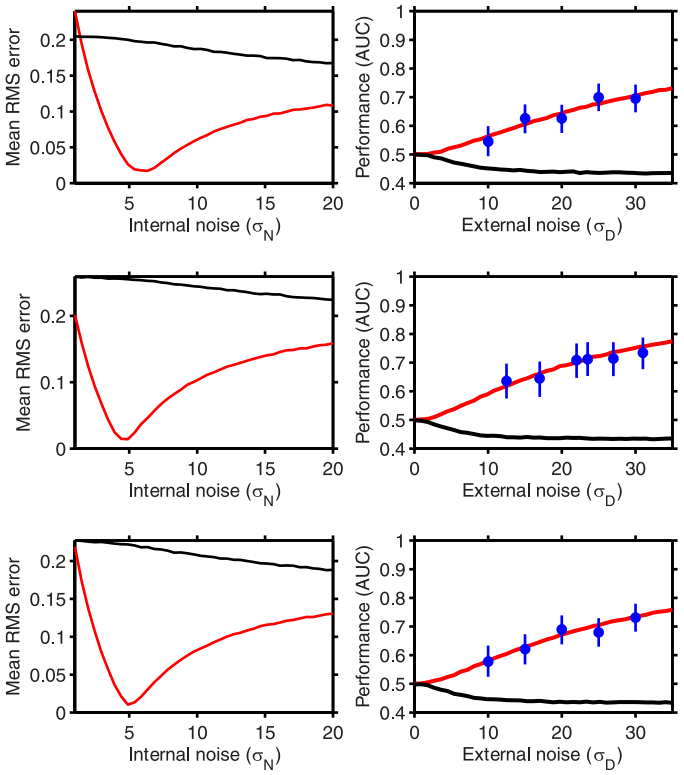

Figure 4. Results for each subject (BV, top; CS, middle; GH, bottom). The left column shows the goodness of fit (cross validation test set error) between predicted and human AUC values. Black lines show MAX-observer goodness of fit, and red lines show MAP-observer goodness of fit. The best fitting (training set) internal noise parameter for each subject was used to calculate predicted performance as a function of external noise for MAP and MAX observers (right column). Human performance is shown by blue data points with 95% confidence intervals.

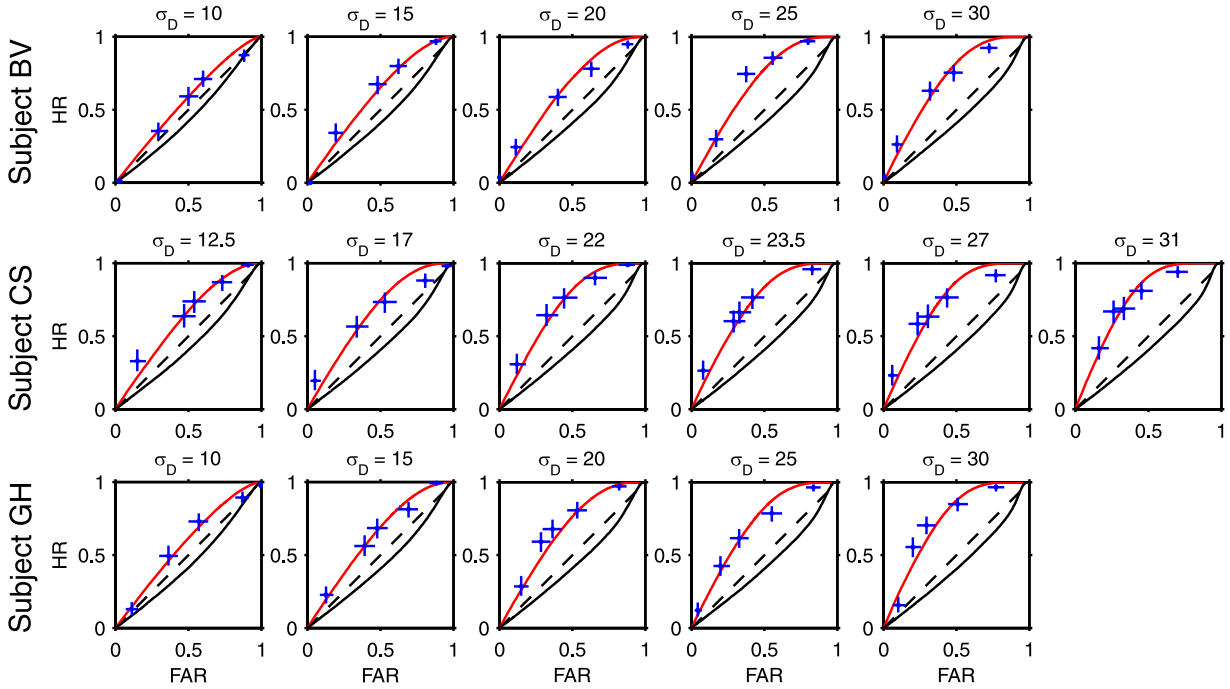

Figure 5. Performance of subjects (blue points) with predicted MAP ROC curve (red) and predicted MAX ROC curve (black). Error bars of FAR/HR points are 95% confidence intervals, estimated using a Dirchlet distribution (see Methods). For each subject, a single internal noise parameter is fitted across all conditions.

MAP observer, values stated above), we plot predicted performance as a function of external noise added to distracters (Figure 4, right column). All subjects (rows of Figure 4) have a clear trend of increasing performance as distracters become more variable (blue data-points; error bars are 95% confidence intervals). The MAP observer predicts this increase in performance (red curves; also see predicted trend in Figure 3). Conversely, the MAX observer not only predicts a performance change in the wrong direction, but also predicts below-chance levels of performance. Due to its limitations, the best the MAX observer can ever do at this particular search task is a performance calculated by area under ROC curve of 0.5 and would require internal noise to be infinite. Even though distracter noise does make distracters more distinguishable from the targets, the MAX observer predicts under chance performance because of the action of the max operation on multiple distracters.

The trends apply to all participants despite their varied level of practice at the task. Qualitatively, it would seem that practice is not required for observers to act in an optimal manner.

It is possible, that the MAP observer could accurately predict AUC performance values, but fail to predict the shape of the ROC curve. As a final test of the MAP observer, we plot each subject's false alarm and hit rate data (Figure 5, blue points with 95% confidence intervals) and compare this to the predicted ROC curve of the MAP observer (red curves). It can be seen that the MAP observer not only captures the area under the ROC curve (Figure 4), but also the shape of the ROC curve itself (Figure 5) across all conditions, for all subjects.

Discussion

Relation to previous SDT work

It is important to clarify how the current results relate to the existing detection theoretic literature applied to visual search. Our findings do not in any way conflict with the previous results. Firstly, the predominant use of the MAX observer in such work will in fact produce identical predictions to the MAP observer in single-feature searches where target and distracter variance is identical (see Figure 1 and Appendix A). In this case, empirical data can be seen as equally consistent with both MAX and MAP observers. In cases which violate these conditions, such as distracter heterogeneity manipulations, then the MAX observer is suboptimal and results in a less convincing, if not poor, explanation of the data (Rosenholtz, 2001).

Shimozaki et al. (2003) have shown that the MAX observer performs suboptimally in a yes/no *cued* detection task when the signal-to-noise ratio or the target's contrast is low. This brings into question the role of a max-of-output mechanism in visual search. Here we have shown that in the distracter variance case, that the MAP observer does fit the empirical data well: we interpret this (alongside

the existing literature) as support for inference processes underlying visual search.

It may be possible to propose a variety of observer hypotheses which still make decisions based on a perceptual dimension but which perform better than the MAX observer. For example, observers based not upon the percepts themselves, but on *differences* between percepts may be able to achieve above MAX level performance. See the *RCref* model (Rosenholtz, 2001) and the *signed-max observer* (Baldassi & Verghese, 2002).

Relation to other models of attentional allocation

Having found poor fits of the MAX-observer to human data, we have claimed that the max-of-outputs type mechanism which acts on a sensory feature dimension should be abandoned (see also Vincent, Troscianko, & Gilchrist, 2007). This can be replaced with a similar max-mechanism, but this operates upon a probability dimension. This has the intuitive appeal that decisions are directly based upon how probable they are to be targets.

This result also has implications for the nature and role of the master map within Guided Search (Wolfe, 1994) and Feature Integration Theory (FIT, Treisman & Gelade, 1980). Within these models perceptual dimensions (or features) are initially processed in independent maps which are then brought together via attention to select the location with the maximum combined activity.

One potential 'work-around' this problem would be to double the complexity of a max-mechanism and suppose that there are 2 thresholds operating on a feature dimension but each with a different polarity. This approach could be described as a MAX-MIN-observer, and could result in predictions akin to the MAP observer. However, such a model would not be as parsimonious and have none of the theoretical elegance that the Bayes-optimal observer has.

Assumptions the brain makes about the world

While the differences between the MAX and the optimal (MAP) observer may seem a very indirect way of understanding the brain, they do, in fact, have meaningful implications about the way the brain makes assumptions about the external world. In viewing natural scenes, it seems reasonable to assume that the entire background (everything excluding the target) can be considered as 'distracters'. *If* the brain made the simplifying assumption that targets and background had equal variance, then in the simplified case of a single feature dimension, the MAX observer would provide the optimal way of detecting targets. We showed the MAX observer provides a poor fit to human data, therefore we argue that the brain probably does not make this simplifying assumption about the world. Instead, we know at the least that the brain incorporates knowledge that the distracters or background can be much more variable than targets.

Because the visual world is complex and consists of many feature dimensions, it could be the case that the brain does make certain assumptions in order to facilitate learning of new target features, or background contexts. Determining precisely what, if any, assumptions about the structure of the environment the brain does use remains to be seen.

Neural basis

Having a simple theoretical framework with which to understand visual search phenomena is appealing, however unless this theory can be realistically implemented by the brain then any theory will remain of limited use. Existing work suggests the simple 3 stage optimal observer (representation, integration and decision) can be readily implemented by the brain. In terms of *representation,* the probabilistic population coding interpretation of neural coding makes convincing arguments that neural tuning curves literally represent probability distributions (Pouget, Dayan, & Zemel, 2000). For example the Gaussian representations of target and distracters (i.e. likelihoods; Figures 3a, 3c, and 3e) discussed in this paper can be implemented by neurons or groups of neurons whose neural tuning curves match the distribution of observed orientations. In terms of *integration,* Gold and Shadlen (2001) and Ma, Beck, Latham, and Pouget (2006) describe how neurons with such tuning curves can be used for probability and likelihood calculations. Whether the brain calculates posterior probability or log posterior odds (see Appendix A) does not affect the fit of theory to data, but further empirical investigation could shed light on neural implementation. In terms of *decision,* the notion that observers make decisions by a threshold mechanism upon sensory dimensions is a fundamental tenant of signal detection theory (for example Ress & Heeger, 2003).

The emerging probabilistic account of visual search

We argue that the Bayesian approach is a natural way to formulate the problem of eye guidance in a noisy and uncertain visual environment, and there is no prior reason to believe the processes underlying visual search should change in a psychophysics setting. Probabilistic accounts of visual search processes are beginning to emerge and are both highly appealing on theoretical grounds, but also very powerful at accounting for empirical observations. For example Torralba, Oliva, Castelhano, and Henderson (2006) showed that humans infer likely locations of targets using

knowledge derived from global scene composition (Oliva & Torralba, 2001). A probabilistic account of accumulating data with internal noise within individual fixations has also been investigated (Carpenter & Williams, 1995). Optimality in visual scanning over multiple fixations has also been examined (Najemnik & Geisler, 2005) with interesting insight into trans-saccadic memory. The optimal feature integration detailed here provides a generative account of where the 'visibility function' in the work of Najemnik and Geisler (2005) may arise from.

A key part of the developing Bayesian probabilistic account of visual search is how knowledge of target features is integrated. Palmer and colleagues first applied signal detection theory to the problem of visual search in 1993 and the approach has been fruitful in accounting for many visual search performance effects such as set size effects, conjunction searches and search asymmetries (see Verghese, 2001, for a review). Until now this approach provided a less than satisfactory account for distracter heterogeneity manipulations; however this obstacle has been removed when realizing features are integrated optimally.

Conclusion

We developed a Bayesian optimal account of feature-based visual attention. Within this account the decision is based on a dimension of probability rather than the more usual (but suboptimal) perceptual dimension, but this alone does not tell us the actual *modus operandi* of humans. In order to empirically determine this, we compared predicted and human performance in a distracter heterogeneity visual search manipulation. Human performance at visual searches exceeded the maximum performance predicted by the SDT account. We take this as evidence that humans do not utilize cues according to the MAX rule. Instead, performance was better accounted for by the Bayesian optimal observer where humans evaluate the probability that each display item is the target given the feature data available: present or absent response is determined by a threshold upon this posterior probability dimension. The contribution of this paper is not so much the formulation of the optimal observer itself (Palmer et al., 2000; Shimozaki et al., 2003), but the empirical data that supports it as a good model of the processes underlying visual search.

Appendix A

Writing the optimal observer using log posterior odds terminology can provide additional insight. In relation to the optimal observer described in the main text: step 1 is identical; step 2 involves calculating the log posterior odds rather than the posterior probability; step 3 is identical, but decision threshold applies to the log posterior odds.

The log posterior odds can be written as below, and we also know the likelihoods are Gaussians and know the priors in relation to N.

$$\log\left(\frac{P(T|u)}{P(D|u)}\right) = \log\left(\frac{P(U_s|T)}{P(U_n|D)}\right) + \log\left(\frac{P(T)}{P(D)}\right). \qquad \text{(A1)}$$

Focusing upon the log likelihood, first term on RHS of Equation A1, we can consider the case where target and distracter uncertainty is identical ($\sigma_T = \sigma_D = 1$) and $\mu_T = \mu_D$. The resulting log likelihood (first term in Equation A1) leaving just the log prior odds (second term in Equation A1), which is a constant $\log[1/(N-1)]$ for a given set size. This can be compared to the flat posterior for the same situation show in Figure 3b.

We also consider the more general situation where target and distracter means are different, $\mu_T = d'$, $\mu_D = 0$ and simplify. This results in a log likelihood of $d'(u - d'/2)$ (also see Wickens, 2002, p. 161) which is also a linear function of the internal percept, u. The log prior odds is just a constant for a given set size, so does not affect the linearity. Because the log posterior odds is a linear function of u, then the display item with the highest internal sensory percept u will also have the highest log posterior odds. In summary, when targets and distracters are Gaussians of equal variance, then the MAX observer is equal to the optimal observer.

Considering the case of targets and distracters with unequal uncertainty ($\sigma_T \neq \sigma_D$), the same simplification procedure of Equation A1 results in quadratic functions of u, regardless of whether the target and distracter means are equal or not. Because of this, the maximum value along the stimulus dimension u will not necessarily equate to the maximum value along the log posterior odds. Thus MAX and MAP observers result in different predictions in the case $\sigma_T \neq \sigma_D$.

Acknowledgments

We wish to thank Richard Newcombe for initial discussions, Elly Martin for help with Appendix A, Ben Tatler for commenting upon earlier versions of the manuscript and Dawn Gray for pilot data. Some of this work was completed while BTV was at University of Bristol, supported by EPSRC grant GR/S47953/01(P). Some of this work has been presented at the International Congress of Psychology, Berlin (Vincent, 2008) and the

European Conference of Visual Perception, Utrecht (Vincent, 2008).

Commercial relationships: none.
Corresponding author: Benjamin T. Vincent.
Email: b.t.vincent@dundee.ac.uk.
Address: School of Psychology, University of Dundee, DD1 4HN, Scotland, UK.

Footnote

[1]Because orientation is a periodic value (i.e. 0° ≡ 360°) then technically, circular normal distributions should be used. However, because we deal with relatively low standard deviations of external noise, Gaussian distributions are entirely satisfactory.

References

Allport, A. (1989). Visual attention. In M. I. Posner (Ed.), *Foundations of cognitive science* (pp. 631–682). Cambridge, MA: MIT Press.

Baldassi, S., & Verghese, P. (2002). Comparing integration rules in visual search. *Journal of Vision, 2*(8):3, 559–570. [PubMed] [Article]

Buswell, G. T. (1935). *How people look at pictures: A study of the psychology of perception in art*. Chicago: University of Chicago Press.

Cameron, E. L., Tai, J. C., Eckstein, M. P., & Carrasco, M. (2004). Signal detection theory applied to three visual search tasks—Identification, yes/no detection and localization. *Spatial Vision, 17,* 295–325. [PubMed]

Carpenter, R. H., & Williams, M. L. (1995). Neural computation of log likelihood in the control of saccadic eye movements. *Nature, 377,* 59–62. [PubMed]

Carrasco, M., & Frieder, K. S. (1997). Cortical magnification neutralizes the eccentricity effect in visual search. *Vision Research, 37,* 63–82. [PubMed]

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review, 96,* 433–458. [PubMed]

Duncan, J., & Humphreys, G. W. (1992). Beyond the search surface. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 578–588. [PubMed]

Eckstein, M. P. (1998). The lower visual search efficiency for conjunctions is due to noise and not serial attentional processing. *Psychological Science, 9,* 111–118.

Eckstein, M. P., Thomas, J. P., Palmer, J., & Shimozaki, S. S. (2000). A signal detection model predicts the effects of set size on visual search accuracy for feature, conjunction, triple conjunction, and disjunction displays. *Perception & Psychophysics, 62,* 425–451. [PubMed]

Gold, J. I., & Shadlen, M. N. (2001). The neural basis of decision making. *Annual Review of Neuroscience, 30,* 535–574. [PubMed]

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Henderson, J. M. (2003). Human gaze control in real-world scene perception. *Trends in Cognitive Sciences, 7,* 498–504. [PubMed]

Land, M. F., & Hayhoe, M. M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Research, 41,* 3559–3565. [PubMed]

Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience, 9,* 1432–1438. [PubMed]

Najemnik, J., & Geisler, W. S. (2005). Optimal eye movement strategies in visual search. *Nature, 434,* 387–391. [PubMed]

Oliva, A., & Torralba, A. (2001). Modelling the shape of the scene: A holistic representation of the spatial envelope. *International Journal in Computer Vision, 42,* 145–175.

Palmer, J. (1994). Set-size effects in visual search: The effect of attention is independent of the stimulus for simple tasks. *Vision Research, 34,* 1703–1721. [PubMed]

Palmer, J. (1998). Attentional effects in visual search: Relating search accuracy and search time. In R. D. Wright (Ed.), *Visual attention*. New York: Oxford University Press.

Palmer, J., Ames, C. T., & Lindsey, D. T. (1993). Measuring the effect of attention on simple visual search. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 108–130. [PubMed]

Palmer, J., Verghese, P., & Pavel, M. (2000). The psychophysics of visual search. *Vision Research, 40,* 1227–1268. [PubMed]

Pashler, H. (1987). Target-distractor discriminability in visual search. *Perception & Psychophysics, 41,* 285–292. [PubMed]

Pouget, A., Dayan, P., & Zemel, R. (2000). Information processing with population codes. *Nature Reviews Neuroscience, 1,* 125–132. [PubMed]

Ress, D., & Heeger, D. J. (2003). Neuronal correlates of perception in early visual cortex. *Nature Neuroscience, 6,* 414–420. [PubMed] [Article]

Rosenholtz, R. (2001). Visual search for orientation among heterogeneous distracters: Experimental results and implications for signal-detection theory models of search. *Journal of Experimental Psychology: Human Perception and Performance, 27,* 985–999.

Shimozaki, S. S., Eckstein, M. P., & Abbey, C. K. (2003). Comparison of two weighted integration models for the cueing task: Linear and likelihood. *Journal of Vision, 3*(3):3, 209–229. [PubMed] [Article]

Tolhurst, D. J., Movshon, J. A., & Dean, A. F. (1983). The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision Research, 23,* 775–785. [PubMed]

Torralba, A. (2003). Modeling global scene factors in attention. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 20,* 1407–1418. [PubMed]

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real world scenes: The role of global features in object search. *Psychological Review, 113,* 766–786. [PubMed]

Treisman, A., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology, 12,* 97–136. [PubMed]

Verghese, P. (2001). Visual search and attention: A signal detection theory approach. *Neuron, 31,* 523–535. [PubMed] [Article]

Verghese, P., & Nakayama, K. (1993). Stimulus discriminability in visual search. *Vision Research, 34,* 2453–2467. [PubMed]

Vincent, B. (2008). Feature-based visual attention: the probabilistic account. *Perception, 37,* 3. [Article]

Vincent, B. T., Troscianko, T., & Gilchrist, I. D. (2007). Investigating a space-variant weighted salience account of visual selection. *Vision Research, 47,* 1809–1820. [PubMed] [Article]

Wickens, T. D. (2002). *Elementary signal detection theory*. New York: Oxford University Press.

Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision, 4*(12):11, 1120–1135. [PubMed] [Article]

Wolfe, J. M. (1994). Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review, 1,* 202–238.

Wolfe, J. M. (1998). What can 1,000,000 trials tell us about visual search? *Psychological Science, 9*.

Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience, 5,* 1–7. [PubMed]

Yarbus, A. L. (1967). *Eye movements and vision*. New York: Plenum.

Zelinsky, G. J., & Sheinberg, D. L. (1997). Eye movements during parallel-serial visual search. *Journal of Experimental Psychology: Human Perception and Performance, 23,* 244–262. [PubMed]