The Theory of Visual Attention without the race: a new model of visual selection

Tobias S. Andersen (ta@imm.dtu.dk)
Informatics and Mathematical Modeling, Technical University of Denmark
2800 Kgs. Lyngby, Denmark

Søren Kyllingsbæk (sk@psy.ku.dk)
Center for Visual Cognition, Department of Psychology, Øster Farimagsgade 2A, 1353 Kbh. K.,
University of Copenhagen, Denmark

Abstract

The Theory of Visual Attention (TVA; Bundesen, 1990) is a comprehensive quantitative account of visual attention, which accounts for many empirical findings and has been extensively applied to clinical studies of attention. According to TVA, perceptual processing of objects occurs in parallel constrained by a limited processing capacity or rate, which is distributed among target and distractor objects with distractor objects receiving a smaller proportion of resources due to attentional filtering. Encoding into a limited visual short-term memory is implemented as a race model. Given its major influence it is surprising that few studies have compared TVA directly to alternative models. Here we insert an algebraically simpler model of encoding into TVA as an alternative to the race model and show that this provides a better fit to Shibuya and Bundesen's (1988) whole and partial report data, which have been a keystone test bed for TVA.

Keywords: Attention; working memory; Theory of Visual Attention; Vision; Psychophysics; Modeling

Introduction

The Theory of Visual Attention (TVA; Bundesen, 1990) incorporates visual perceptual processing, attentional filtering and encoding into visual short-term memory (VSTM) in a unified quantitative model. The model has been extended to account for results from a wide variety of experimental paradigms (Logan, 1996; Logan & Gordon, 2001), and the neural implementation of TVA (NTVA) has been applied to results from single cell studies (Bundesen, Habekost, & Kyllingsbæk, 2005). Despite the extensiveness of the TVA based theoretical framework, we are aware of only a few recent studies (Dyrholm, Kyllingsbæk, Espeseth, & Bundesen, 2011; Kyllingsbaek, Markussen, & Bundesen, 2011; Petersen & Andersen, 2012) challenging the specific details of the model using standard model assessment methods. Of these studies we will include Petersen and Andersen's (2012) findings that the log-logistic psychometric function inserted into TVA leads to improved performance in the current study.

Computational models of cognition such as TVA offer both theoretical and practical advantages. The theoretical advantages include the strict quantitative formulation of cognitive modules, the definition of which can otherwise prove to be elusive. Computational models can also be applied to a range of experimental paradigms and help arrive at a unified interpretation. This can be of practical use as the assessment of the function of cognitive modules is of great importance in clinical psychology and neuro-pharmacology. In this vein, TVA has been extensively applied to studies of clinical populations (Habekost & Starrfelt, 2009) and to the effect of psychoactive drugs (Finke, et al., 2010; Vangkilde, Bundesen, & Coull, 2011). Many of these studies base their assessment on estimates of the parameters in TVA and therefore rely on TVA precisely reflecting the actual computational mechanisms underlying visual attention. This makes it the more pressing to assure that this is indeed the case by comparing the specifics of TVA to competing models.

Whole and partial report tasks have been a keystone test bed for TVA. In whole report tasks, a number of objects, typically letters or digits, are presented to the observer. The task of the observer is to identify and report the objects presented. The exposure duration is typically brief (<200 ms) in order to avoid eye movements so that the information available can be assumed to be near constant across the stimuli and throughout the stimulus duration. Partial report tasks are like whole report except that in addition to the target

objects, a number of distractor objects are also presented. Some characteristic, like color, location or object category (e.g. letters vs. digits) distinguishes targets from distractors. The task of the observer is to report only the target objects and ignore the distractors.

Performance in whole report tasks is limited by perception and memory. In order for the target objects to be correctly reported, they must be perceived. This depends on stimulus attributes such as contrast, exposure duration, size, complexity and the number of stimulus categories (Pelli, Burns, Farell, & Moore-Page, 2006). Since these limitations exist also when only a single object is present the effect of these stimulus attributes can be studied in single letter identification experiments (Petersen & Andersen, 2012).

When multiple objects are presented the single letter psychometric functions cannot explain performance. Instead, the psychometric function needs to be adjusted. In TVA the adjustment is based on the assumption that the sum of processing resources, defined as the sum of hazard rates, is constant (Shibuya & Bundesen, 1988).

In partial report tasks, performance depends also on the ability to filter out the irrelevant distractor objects through selective attention in order to avoid their interference with perceptual processing and their taking up working memory capacity. If filtering is perfect, performance in partial report tasks should match that of whole report tasks with the same number of target objects. Shibuya and Bundesen (1988) showed that this is not the case and that the filtering process is imperfect. TVA models filtering as a smaller amount of processing resources being allocated to distractor objects.

Even when contrast and exposure duration are more than sufficient for all letters to be correctly identified according to the adjusted psychometric functions, observers fail to base their report on more than about four objects (Sperling, 1960). This seems to be due to limitations on VSTM rather than on perception *per se*. In TVA the mechanism of encoding is a race, so that objects are encoded into VSTM when they are perceptually processed but only if VSTM capacity is still available, i.e. if it has not already been occupied by other objects.

TVA is thus able to describe performance in whole and partial report tasks with a given number of targets and distractors based on performance in single object identification in the form of the psychometric function. It does this based on assumptions of how multiple targets affect perceptual processing, the process of filtering and encoding into a limited VSTM. We find it difficult to envision a model that would not partition visual perception, attention and short-term memory into these components as does TVA but we find that there is room to examine the specific implementation of these stages.

In the following we shall examine the encoding stage of TVA, the race model. We will insert a different model of the encoding stage into TVA and compare the two encoding models' abilities to describe Shibuya and Bundesen's (1988) whole and partial report data. We will do this using either the exponential psychometric function conventionally used in TVA or the log-logistic function that Petersen and Andersen (2012) found to improve performance.

Methods

Modeling

The psychometric function and distributing resources

In TVA, perceptual processing of a single object is typically described by the exponential psychometric function

$$F(t) = 1 - \exp\left(-v_t\left(t - t_0\right)\right), t > t_0$$
$$F(t) = 0 \qquad\qquad\qquad\qquad , t > t_0$$

where F is the probability of correctly identifying the object, v_t is the rate of processing for the target object, t is the exposure duration and t_0 is a short time interval between stimulus onset and the beginning of perceptual processing. In terms of probability theory, the rate, v_t, is the hazard rate and $v_t(t\text{-}t_0)$ is the cumulative hazard rate, the hazard rate integrated over time. When only a

single target is presented the sum of processing resources, or hazard rates, C, is allocated to that target so that $\nu_t = C$. In whole report, when multiple targets are presented, the objects are typically arranged at equal distances from the fixation point so that it is reasonable to assume that they receive equal shares of the processing resources, i.e. $\nu_t = C/T$, where T is the number of targets. In partial report, distractor objects are assumed to receive a proportionally smaller share of processing resources due to attentional filtering so that $\nu_d = \alpha\nu_t$. From this, we can deduce that $\nu_t = C / (T+\alpha D)$ where D is the number of distractors (Bundesen, 1990).

In a recent study Petersen and Andersen (2012) showed that other psychometric functions can be inserted into TVA and that this, in general, improves the performance of the model. The log-logistic function gave the best fit of those functions having two free parameters like the exponential function. Therefore we will use it here. The log-logistic can be expressed as

$$F(t) = \frac{1}{1+\left(\frac{t}{t_0}\right)^{-\nu_t}}$$

Although the parameters t_0 and ν_t describe the shift and the slope of the psychometric function respectively just as for the exponential function, their exact meaning is different than for the exponential function. The shift, t_0, is here the 50% correct threshold. Unlike the exponential function, the hazard rate is not explicit in the expression for the log-logistic function but the cumulative hazard rate, Λ_t, can be derived to be

$$\Lambda_t = -\log(1-F) = \log\left(1+\left(\frac{t}{t_0}\right)^{\nu_t}\right)$$

Distributing processing resources according to TVA with the log-logistic function becomes simpler if we notice that the assumption of a constant sum of hazard rates is equivalent to a constant sum of cumulative hazard rates. When only a single object is presented the cumulative hazard rate is thus $\Lambda_t = C_{cum}$. From this the response probabilities in whole and partial report can be calculated by setting the cumulative hazard rate to $C_{cum} / (T+\alpha D)$.

Encoding into a limited VSTM

The previous section outlined TVA applied to the case of whole and partial report when the total number of objects does not exceed the capacity of VSTM. In that case we can calculate the probability of the score, j, which is the number of correctly reported target objects, as

$$P(j) = \binom{T}{j}[F(t)]^j[1-F(t)]^{T-j}$$

This expression is derived from the binomial distribution giving the probability of encoding j targets. The number of encoded target objects is termed the *score*.

When the number of objects exceeds VSTM capacity selection of the objects to encode is needed. According to TVA the selection happens as a race for free slots in VSTM; a race that ends when all slots are occupied or when perceptual processing ends. Inserting the race model into TVA is somewhat algebraically complex but allows calculating the score probability, i.e. the probability of correctly reporting a certain number of target objects. Detailed expressions and derivations are given in Petersen and Andersen (2012).

Here we introduce a different model of selection of objects to be encoded by conditioning on the total number of objects encoded being no greater than VSTM capacity, i.e. $j+m \leq K$ where m is the number of distractor objects encoded. This probability is calculated by calculating the score probabilities for $j = 1,...,T$ and $m \leq \min(D,K-j)$

$$P(j) = \binom{T}{j}[F(t)]^j[1-F(t)]^{T-j} \times \sum_{m=0}^{\min(D,K-j)} \binom{D}{m}[G(t)]^m[1-G(t)]^{D-m}$$

Conditioning on $j+m \leq K$ is then implemented by normalization of the probability mass function $P(j)$. Here, the psychometric function for distractor objects is denoted $G(t)$. Note that the number of encoded distractor objects, m, is considered an unobservable nuisance parameter, which is summed out.

For both encoding models, VSTM capacity, K, is allowed to take non-integer values, which are implemented as a mixture model where the VSTM capacity is the ceiling value of K, $\lceil K \rceil$, with a probability of $\mathrm{mod}(K, \lfloor K \rfloor)$ where $\lfloor K \rfloor$ is the floor value of K and $\lfloor K \rfloor$ with a probability of $1 - \mathrm{mod}(K, \lfloor K \rfloor)$.

Model evaluation

As testing ground for comparing the two models of encoding we choose Shibuya and Bundesen's (1988) whole and partial report data that have been influential in the development of TVA (Bundesen, 1990). The data set consists of score counts for two observers each performing 6,480 trials with varying number of target and distractor elements and exposure durations. The observers were instructed to report the identity of targets only when they were reasonably confident in order to minimize the effect of guessing.

Only very rarely did the observers achieve scores greater than 4. Following the example of Bundesen (1990) we have registered these responses as scores of 4. The encoding models can be extended to account for these higher scores by allowing the VSTM capacity to vary between three integer values rather than just two but this requires an additional free parameter, which is difficult to justify by the ability to model only few of thousands of trials.

Results

Table 1 displays the goodness of fits in terms of the negative logarithm of the likelihood for the two models of encoding and the two psychometric functions fitted to both observers in Shibuya and Bundesen's (1988) data. Note that the encoding models and psychometric functions have the same number of free parameters.

The goodness of fits in Table 1 confirms that the log-logistic psychometric function provides a better fit than the exponential psychometric function as found by Petersen and Andersen (2012) and also that the conditioning model offers an additional, although slight, improvement in the goodness of fit.

Table 1: Goodness-of-fits

	Selection model	
Psychometric function	Race	Conditioning
Exponential	1579	1552
Log-logistic	1331	1273

To further examine the fits of the encoding models Figure 1 displays the cumulative score proportions, i.e. the proportion of responses to a given stimulus type with at least j correctly reported targets along with model fits for both encoding models with the log-logistic psychometric function for subject HV. As is evident from Figure 1, the model fits are very similar. It takes careful inspection to see that there are, in fact, systematic differences. The clearest difference is that when six targets are presented both encoding models tend to overestimate the cumulative score proportion but the conditioning model less so than the race model. Also, when the number of distractors is no greater than two, both models tend to underestimate the cumulative score proportion for exposure durations between 30–70 ms but the conditioning model less so.

For the briefest exposure durations of 10 ms observers rarely reported any targets. In Bundesen's (1990) analysis the few trials in which they did were discarded so that the score was assumed to be zero. This might favor the exponential psychometric function as it constrains the score to be zero for exposure durations shorter than t_0. We therefore fitted the models to the data with this data adjustment. The conditional model still fitted the data better but more so with the exponential psychometric function than with the log-logistic.

Table 2 lists the parameter values for the fits. Note that the VSTM capacity, K, and filter-parameter, α, are comparable across both encoding model and psychometric function. They seem, however, to vary very little with these model variations within observers. The temporal threshold, t_0, and processing capacity, C, are not comparable across psychometric functions, only across encoding models. The temporal threshold seems also to vary very little with encoding model within observer and psychometric function. For the log-logistic function the processing capacity, i.e. the sum of hazard rates, varies over time and is therefore given for $t = t_0$. The processing capacity is slightly, but consistently, greater for the conditioning model than for the race model. This difference may seem slightly more pronounced for the log-logistic psychometric function (4 s^{-1} averaged over the two observers) than for the exponential function (2 s^{-1} averaged over the two observers) but this might be due to the difference in magnitude of C as the relative differences were similar (7% for the log-logistic model and 5% for the exponential model averaged over the two observers).

Table 2: Estimated parameters for psychometric functions (Psy. F.), observer (Obs), and model.

Psy. F.	Obs.	Model	K	α	C	t_0
Log-log.	MP	Race	3.8	0.40	57	0.036
		Cond.	3.9	0.41	61	0.036
	HV	Race	3.3	0.56	50	0.033
		Cond.	3.2	0.52	53	0.034
Exp.	MP	Race	3.9	0.39	37	0.010
		Cond.	4.0	0.38	38	0.010
	HV	Race	3.3	0.55	35	0.010
		Cond.	3.2	0.52	37	0.010

Discussion

The differences that we found between the fits of the encoding models are consistent. They are, however, also small. This warrants care in model selection and this is, in fact, our main point. The

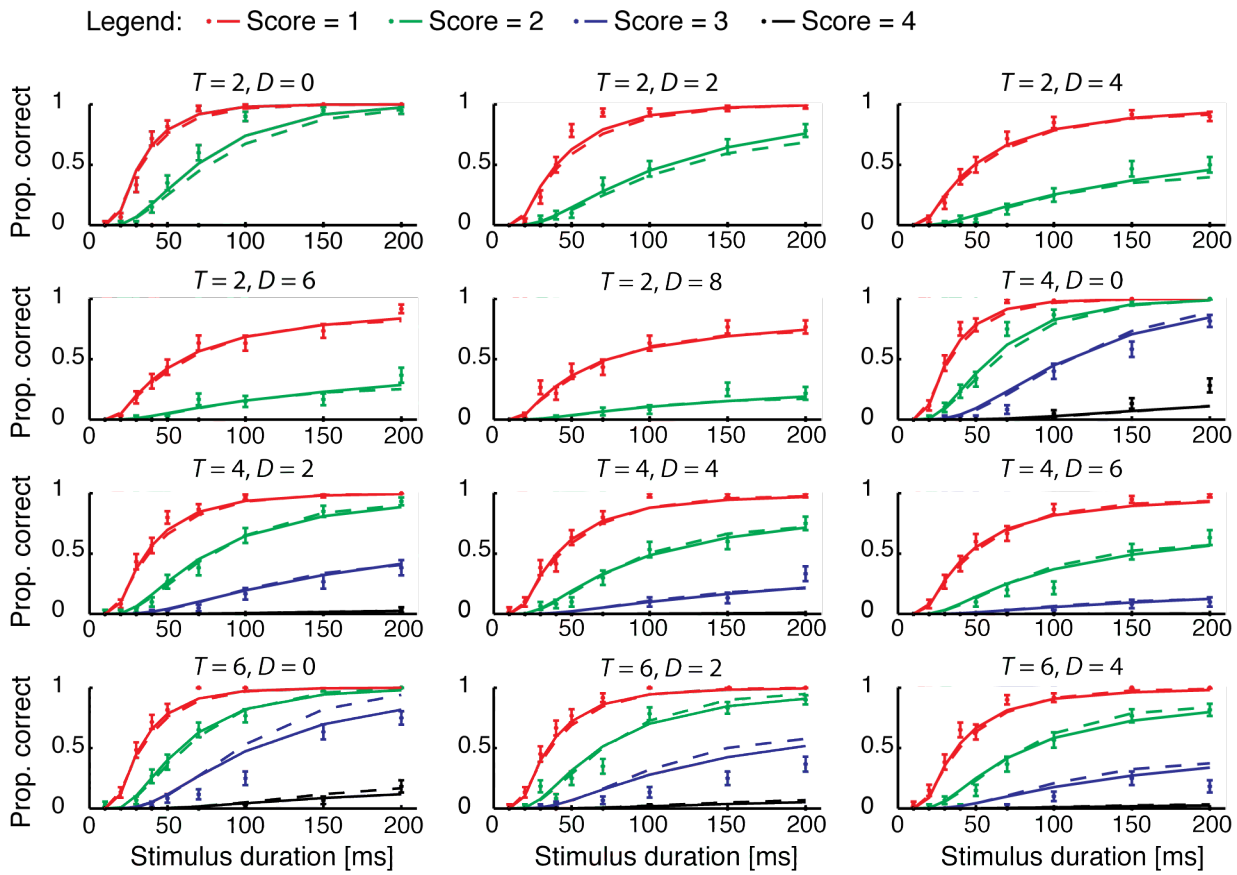

Figure 1: Cumulative score proportions for Shibuya and Bundesen's (1988) whole and partial report experiment with estimates from the conditional (solid line) and race (dashed line) encoding models using the log-logistic psychometric function. T and D at the top of each graph indicate the number of presented target and distractor objects respectively.

model fits do not provide strong evidence in favor of one model of selection over another. We find that this is a strong point as the race model has remained unchallenged as the model of selection for two decades of TVA based research.

Parameter estimates varied very little with the type of psychometric function and encoding model. The only consistent differences in parameter estimates between the two encoding models were in the processing capacity, *C*. This difference should however be compared to the variability within observers estimated by Finke et. al. (2005) to be as high as 20% although variability between observers can also be as little as 10% (Vangkilde, et al., 2011), which is far less than the differences observed between clinic populations and normal controls (Starrfelt, Habekost, & Leff, 2009; Vangkilde, et al., 2011) yet greater than the differences between the models tested here. We therefore preliminarily conclude that parameter estimation is robust to variations in the type of psychometric function and encoding model with the caveat that studies of greater populations than the two observers studied may reveal greater variability.

Model comparison should be based on the models' ability to describe the data, here given by the goodness-of-fit; model flexibility, here given by the number of free parameters; but also on model interpretability. The interpretation of the race model is straightforward; it explicitly gives a mechanism for selection of objects to be encoded into VSTM. On this point the conditioning model is vague. We do not understand what mechanism, cognitive or neural, that could implement selection by conditioning but find that this is an interesting topic for future studies.

Acknowledgments

The authors thank Claus Bundesen and Hitomi Shibuya for making the data from their whole and partial experiment (Shibuya & Bundesen, 1988) available.

References

Bundesen, C. (1990). A Theory of Visual Attention. *Psychological Review, 97*(4), 523–547.

Bundesen, C., Habekost, T., & Kyllingsbæk, S. (2005). A Neural Theory of Visual Attention: Bridging Cognition and Neurophysiology. *Psychological Review, 112*(2), 291–328.

Dyrholm, M., Kyllingsbæk, S., Espeseth, T., & Bundesen, C. (2011). Generalizing parametric models by introducing trial-by-trial parameter variability: The case of TVA. [doi: 10.1016/j.jmp.2011.08.005]. *Journal of Mathematical Psychology, 55*(6), 416-429.

Finke, K., Dodds, C. M., Bublak, P., Regenthal, R., Baumann, F., Manly, T., et al. (2010). Effects of modafinil and methylphenidate on visual attention capacity: a TVA-based study. *Psychopharmacology (Berl), 210*(3), 317-329.

Habekost, T., & Starrfelt, R. (2009). Visual attention capacity: a review of TVA-based patient studies. *Scand J Psychol, 50*(1), 23-32.

Kyllingsbaek, S., Markussen, B., & Bundesen, C. (2011). Testing a poisson counter model for visual identification of briefly presented, mutually confusable single stimuli in pure accuracy tasks. *J Exp Psychol Hum Percept Perform*.

Logan, G. D. (1996). The CODE theory of visual attention: an integration of space-based and object-based attention. *Psychol Rev, 103*(4), 603-649.

Logan, G. D., & Gordon, R. D. (2001). Executive control of visual attention in dual-task situations. *Psychol Rev, 108*(2), 393-434.

Pelli, D. G., Burns, C. W., Farell, B., & Moore-Page, D. C. (2006). Feature detection and letter identification. *Vision Res, 46*(28), 4646–4674.

Petersen, A., & Andersen, T. S. (2012). The effect of exposure duration on visual character identification in single, whole and partial report. *Journal of Experimental Psychology: Human Performance and Perception, In Press*.

Shibuya, H., & Bundesen, C. (1988). Visual Selection from Multielement Displays: Measuring and Modeling Effects of Exposure Duration. *Journal of Experimental Psychology Human Perception Performance, 14*(4), 591–600.

Sperling, G. (1960). The Information Available in Brief Visual Presentions. *Psychological Monographs, 74*(11), 1-29.

Starrfelt, R., Habekost, T., & Leff, A. P. (2009). Too little, too late: reduced visual span and speed characterize pure alexia. *Cereb Cortex, 19*(12), 2880-2890.

Vangkilde, S., Bundesen, C., & Coull, J. T. (2011). Prompt but inefficient: nicotine differentially modulates discrete components of attention. *Psychopharmacology (Berl), 218*(4), 667-680.