



Modeling Visual Cognition

– encoding, attention, and short-term memory

Doctoral Dissertation in Psychology

Søren Kyllingsbæk

Department of Psychology
University of Copenhagen

Copyright © 2013 by Søren Kyllingsbæk

Typeset by Søren Kyllingsbæk

Printed in Denmark at Grafisk – Københavns Universitet, Copenhagen, Denmark

ISBN 978-87-996351-1-5

Philosophy is written in this grand book—
I mean the universe—which stands continually open to our gaze, but it cannot be understood unless one first learns to comprehend the language and interpret the characters in which it is written. It is written in the language of mathematics, [...] without these, one is wandering around in a dark labyrinth.

Galileo

Articles

1. Kyllingsbæk, S., Markussen, B., & Bundesen, C. (2012). Testing a Poisson counter model for visual identification of briefly presented, mutually confusable single stimuli in pure accuracy tasks. *Journal of Experimental Psychology: Human Perception and Performance*, *38*, 628-642.
2. Kyllingsbæk, S., & Bundesen, C. (2007). Parallel processing in a multifeature whole-report paradigm. *Journal of Experimental Psychology: Human Perception and Performance*, *33*, 64-82.
3. Kyllingsbæk, S., Valla, C., Vanrie, J., & Bundesen, C. (2007). Effects of spatial separation between stimuli in whole report from brief visual displays. *Perception & Psychophysics*, *69*, 1040-1050.
4. Bundesen, C., Habekost, T., & Kyllingsbæk, S. (2005). A neural theory of visual attention. Bridging cognition and neurophysiology. *Psychological Review*, *112*, 291-328.
5. Petersen, A. E., Kyllingsbæk, S., & Bundesen, C. (2012). Measuring and modeling attentional dwell time. *Psychonomic Bulletin & Review*, *19*, 1029-1046.

6. Kyllingsbæk, S. (2006). Modeling visual attention. *Behavior Research Methods*, *38*, 123-133.
7. Kyllingsbæk, S., Sy, J. L., & Giesbrecht, B. (2011). Understanding the allocation of attention when faced with varying perceptual load in partial report: A computational approach. *Neuropsychologia*, *49*, 1487-1497.
8. Kyllingsbæk, S., & Bundesen, C. (2009). Changing change detection: Improving the reliability of measures of visual short-term memory capacity. *Psychonomic Bulletin & Review*, *16*, 1000-1010.
9. Dyrholm, M., Kyllingsbæk, S., Espeseth, T., & Bundesen, C. (2011). Generalizing parametric models by introducing trial-by-trial parameter variability: the case of TVA. *Journal of Mathematical Psychology*, *55*, 416-429.

Note: The experimental data presented in articles 2 and 3 have been presented in a preliminary form in the author's PhD thesis published in 2001 (Kyllingsbæk, 2001).

Preface

This doctoral dissertation represents a collection of experimental and theoretical studies by the author accomplished in collaboration with colleagues, PhD students, and students at the Center for Visual Cognition, Department of Psychology, at the University of Copenhagen as well as colleagues at other departments at the University of Copenhagen and collaborators at the University of California, Santa Barbara, and the University of Oslo. The research was supported by the Carlsberg Foundation, the Danish Council for Strategic Research, the Sapere Aude Program of the Danish Research Council for Independent Research, and the University of Copenhagen.

I would like to thank Claus Bundesen for more than 20 years of inspiring collaboration and my co-authors Mads Dyrholm, Thomas Espeseth, Thomas Habekost, Bo Markussen, Anders E. Petersen, Jocelyn L. Sy, Christian Valla, and Jan Vanrie. A special thank go to Barry Giesbrecht for inviting me into his lab during my sabbatical at the UCSB. I would also like to thank all my colleagues and the students at the Center for Visual Cognition for making all the days in the lab enjoyable and Jessica Bahr England for revising the manuscript. Finally, I would like to thank my family, Hanne, Jeppe, and Asger for their continued support.

Contents

1	Outline	1
2	Encoding of Visual Information	2
3	A Neural Theory of Visual Attention	8
4	NTVA and other Theories	22
5	Visual Memory Systems	31
6	An Integrated View of Visual Cognition	44
7	Danish Summary	47
8	References	48

1 Outline

The topic of the research presented in this dissertation is visual cognition in general, and more specifically the development of computational models of visual perception, visual attention, and visual short-term memory (VSTM). The models were constructed with the aim of being as parsimonious as possible, while exhibiting high validity in relation to both their structure and parameters. The latter determines whether the models are valid in terms of their direct link to psychologically meaningful concepts such as the rate of encoding of stimulus information and the amount of information that can be retained in memory. An additional aim of the research presented in this dissertation was to extend these models to explain the mechanism of the underlying neural structures, which form the biological basis of the cognitive processes.

The first part of the dissertation comprises a presentation of experimental studies and theoretical work on the question of how visual information is encoded when one or more stimuli are identified. This leads to a presentation of the Neural Theory of Visual Attention (NTVA; Bundesen, Habekost, & Kyllingsbæk, 2005, 2011) followed by an integrated view of visual memory where the nature of and links between sensory memory, visual short-term memory, and visual long-term memory are discussed.

2 Encoding of Visual Information

Visual identification of stimuli is achieved through a series of processing steps, beginning with early sensory processing in the retina and continuing in the lateral geniculate nuclei of the thalamus and early visual areas in the occipital cortex (e.g., Wandell, 1995). During these early stages, visual object segmentation is performed whereby the visual scene is partitioned into candidate regions that correspond to distinct objects in the world (e.g., Logan, 1996; Marr, 1982). Eventually, the information from these segmented areas makes contact with representations of different visual categories/features in visual long-term memory (VLTm). In the present framework, it is assumed that this process of matching the stimulus information with representations in VLTm leads to computation of *sensory evidence values*, $\eta(x, i)$. These values specify to what degree a given object x has a certain visual feature i . The sensory evidence is then weighted by the subjective *perceptual response bias*, β_i , associated with each visual category i , so that expectancy and utility may influence the resulting categorization of the stimulus. Theories of the perceptual decision process that incorporate the basic notion of sensory evidence and perceptual response bias were proposed early on by Shepard (1957, 1958a, 1958b) and Luce (1963) in the famous Shepard-Luce bias-choice model.

After the visual stimuli have been categorized, this information must be retained in order to influence sub-

sequent behavior. A visual short-term memory (VSTM) system is responsible for this. The capacity of VSTM is severely limited compared to VLTM and is only thought to hold information from about four items (e.g., Shibuya & Bundesen, 1988; Sperling, 1960). Note however that when one or two stimuli are presented, the limitation in VSTM capacity has no consequences if one assumes that the effect of forgetting is negligible. In a series of experiments, we have investigated and modeled these visual encoding and perceptual decision processes.

2.1 Single Stimulus Identification

In Kyllingsbæk, Markussen, and Bundesen (2012), we presented participants with single stimuli (digits and Landolt’s rings) for identification. Presentations were brief (≤ 100 ms) and post masked. We used a non-speeded response procedure in which participants were told that their reaction time would *not* be measured and they would be able to use as much time as needed for making their decision. We found that a *Poisson Counter Model* could accurately explain the encoding process. The model implies that during stimulus analysis, tentative categorizations that stimulus i belongs to category j are made at a constant Poisson rate, $v(i, j)$. The accumulation of information is continued until the stimulus disappears, and the overt response is based on the categorization made the greatest number of times. In other words, when presented with a brief visual stimulus, participants will accumulate evidence of different

possible categorizations of the stimulus while it is being exposed. After the offset of the stimulus, the participant will reach a decision by comparing the evidence and choosing the response for which most evidence has been accumulated. Given these assumptions, the probability, $P(i, j)$, of reporting category j when presented with stimulus i and the counts accumulated for category j is higher than the counts for any other categories, may be derived:

$$P(i, j) = \sum_{n=1}^{\infty} \frac{v(i, j)^n (t - t_0)^n}{n!} e^{-v(i, j)(t - t_0)} \quad (1)$$

$$\cdot \prod_{k \in R - \{j\}} \sum_{m=0}^{n-1} \frac{v(i, k)^m (t - t_0)^m}{m!} e^{-v(i, k)(t - t_0)},$$

where n represents the number of counts for categorization j , R is the set of responses, k runs across the set of responses excluding j , m represents the number of counts for these responses, and t_0 is the perceptual threshold so that $P(i, j) = 0$ for $t_0 > t$.

Similarly, equations for the probability of reporting when ties between several counters occur or when none of the categories accumulate any counts may also be derived (see Kyllingsbæk et al., 2012).

2.2 Identification of Multiple Stimuli

How do we acquire information when several visual stimuli are presented for identification? Most theories of visual cognition state that features from a single object may be encoded in parallel, a claim supported by seminal work within neurophysiology and neuroimaging (e.g., Zeki, 1993; Bartels & Zeki, 2004). However, when several visual stimuli must be identified, many influential theories claim that encoding happens in series, with one object encoded at a time (e.g., Treisman, 1988; Wolfe, 1994). In Kyllingsbæk and Bundesen (2007), we modeled data from a novel *multi-feature whole report paradigm* in which the participant had to report several features from simultaneously presented visual stimuli. In one of these experiments, two colored bars were presented and the task of the participant was to report both the color and the orientation of the two stimuli. We found strong evidence for parallel independent processing in this and several other experiments using different combinations of visual features. Specifically, stochastic independence was found between features encoded from the same object and between features from different objects. Thus when we encode information from several stimuli in the visual field, the visual features are processed in parallel and independently of each other.

2.3 Identification and Visual Crowding

The visual system has a limited number of processing resources, which must be allocated optimally when we encode visual stimuli. In Kyllingsbæk, Valla, Vanrie, and Bundesen (2007), we manipulated the spatial separation between several stimulus letters in whole report while keeping the eccentricity of the stimuli constant. We found that the proportion of correctly identified letters was a strictly increasing decelerating function of the spatial separation of the stimuli.

To explain the results, we developed a computational model of visual crowding (i.e. lateral masking, e.g. Bouma, 1970). The model assumes that visual processing resources may be defined in terms of dynamic remapping of receptive fields of neurons in the visual cortex. When several stimuli are presented within the receptive field of a neuron, the stimuli compete for representation, resulting in only one of the stimuli being represented by one neuron at any given time. When the spatial separation between stimuli is large, very little competition takes place and visual processing resources (neurons) are independently allocated to process the stimuli. In contrast, when the spatial separation is small, many neurons will have several stimuli within their receptive field. This results in competition within many neurons and strong dependencies in the allocation of visual processing resources between stimuli. The model enabled us to estimate the size of the receptive fields involved in the processing of the letter

stimuli. The estimated radius of the receptive fields was 29 degrees of visual angle, which corresponded closely with neurophysiological findings of recordings in the inferotemporal cortical visual area of macaque monkeys (see Desimone & Gross, 1979; Desimone & Ungerleider, 1989).

In conclusion, three basic properties of visual encoding were found: (a) Poisson based accumulation of information, (b) parallel independent processing both within and between visual objects, and (c) allocation of processing resources understood as dynamic remapping of cortical receptive fields. These properties form a significant theoretical elaboration that links closely to our development of a neural implementation of Bundesen's (1990) Theory of Visual Attention, which I will describe in the following section.

3 A Neural Theory of Visual Attention

3.1 Background

The Neural Theory of Visual Attention (NTVA) presented in Bundesen et al. (2005) is based on the earlier Theory of Visual Attention (TVA) by Bundesen (1990). TVA is a combined theory of recognition and selection. Whereas many theories of visual attention separate the two processes both in time and in representation, TVA instantiates the two processes in a unified mechanism implemented as a race model of both selection and recognition. In other words, when an object in the visual field is recognized, it is also selected at the same time and vice versa. By the unification of selection and recognition TVA tries to resolve the long-standing debate of early versus late selection. The first position claims that selection occurs prior to recognition (e.g., Broadbent, 1958) while the second claims that recognition is the precursor for selection (e.g., Deutsch & Deutsch, 1963). According to TVA, elements in the visual field are processed in parallel. Visual processing is a two-stage process comprised of (a) an initial match of the visual impression with VLTm representations followed by (b) a selection/recognition race for representation in VSTM.

3.2 The first stage: computation of η values

According to TVA, visual processing starts with a massively parallel comparison (matching) between objects in the visual field and representations in VLTM. This process has unlimited capacity in the sense that the time it takes to process objects is independent of the number of objects in the visual field. The end result of the matching process is the computation of sensory evidence values, $\eta(x, i)$, each measuring the degree of a match between a given object x and a long-term memory representation (category) i . η values are affected by the visibility (e.g. contrast) of the visual stimuli as well as the degrees of pattern match between the stimuli and the representations in VLTM. The latter is affected by learning, which may lead to change in or even development of new representations in VLTM. For example, one may learn to read one's own name faster than other first names (cf. Bundesen, Kyllingsbæk, Houmann, & Jensen, 1997). Altogether, η values are affected only by “objective” properties of the visual environment and the VLTM of the perceiver, not by “subjective” properties such as the pertinence of certain objects or categorization bias.

3.3 The second stage: the race

Different categorizations of the objects in the visual field compete for entrance into VSTM in a stochastic race process. The capacity of VSTM is limited to K el-

ements, typically around four. Categorizations of the first K visual objects to finish processing are stored in VSTM (the first K winners of the race). Categorizations from other elements are lost. Note that categorizations from elements already represented by other categorizations may freely enter VSTM even though it is full. Thus VSTM is limited with respect to the number of elements of which categorizations may be stored, not the number of categorizations of the elements represented (see Awh, Barton, & Vogel, 2007; Duncan, 1984; Luck & Vogel, 1997, but see also Wheeler & Treisman, 2002)

As stated above, visual objects are processed in parallel in TVA. Furthermore, the theory assumes independence between visual categorizations of different objects and between different types of categorizations of the same object. Consider a situation in which only two objects are present in the visual field and are judged according to color and shape. Assuming that VSTM capacity is larger than two, VSTM is not a limiting factor. If attentional parameters are kept constant, TVA predicts that the probability of the first object being correctly categorized with respect to color is independent of whether the object is correctly categorized with respect to shape. The probability is also independent of whether the second object is correctly categorized with respect to color or shape (as we found in Kyllingsbæk & Bundesen, 2007).

To determine the rate of processing of each categorization of an element, the η values are combined with

two types of “subjective” values, pertinence and bias. As suggested by Broadbent (1971), two different attentional mechanisms are necessary for adequate behavior: one for *filtering* (based on pertinence) and one for *pigeonholing* (based on bias). For example, if participants are instructed to report the identity of black target letters amongst white distractor letters, the white distractors must be filtered out, and the black targets must be categorized with respect to letter identity. In TVA terms, pertinence should be high for black stimuli and low for white stimuli and bias should be high for letter identities and low for all other categories. In other words, filtering is based on color and pigeonholing on the letter identity. The rate of processing $v(x, i)$ of a categorization in the race is given by two equations. By the *rate equation*,

$$v(x, i) = \eta(x, i)\beta_i \frac{w_x}{\sum_{z \in S} w_z}, \quad (2)$$

where $\eta(x, i)$ is the strength of the sensory evidence that element x belongs to category i , β_i is the perceptual bias associated with i , S is the set of elements in the visual field, and w_x and w_z are attentional weights for elements x and z . Thus the rate of processing is determined by the strength of the sensory evidence that object x is of category i , weighted by the bias towards making categorizations of type i , and by the relative attentional weight of object x (given by the ratio of w_x over the sum of the attentional weights of all objects in the vi-

sual field). The attentional weights are in turn given by the *weight equation*,

$$w_x = \sum_{j \in R} \eta(x, j) \pi_j, \quad (3)$$

where R is the set of perceptual categories, $\eta(x, j)$ is the strength of the sensory evidence that element x belongs to category j , and π_j is the pertinence (priority) value associated with category j . The distribution of pertinence values defines the selection criteria at any given point in time (filtering). By Equation 3, the attentional weight of object x is a weighted sum of pertinence values, where each pertinence value π_j is weighted by the degree of evidence that object x is actually a member of category j (see also Nordfang, Dyrholm, & Bundesen, 2012).

When η , β , and π values are assumed to be constant, processing rates are also constant. Consequently, the distribution of processing times will be exponential (cf. Bundesen, 1990). Both Bundesen and Harms (1999) and Shibuya and Bundesen (1988) found strong evidence that processing times may be exponential to a good approximation in single stimulus identification and in whole and partial report, respectively (but see also Dyrholm, Kyllingsbæk, Espeseth, & Bundesen, 2011; Petersen & Andersen, 2012; Petersen, Kyllingsbæk, & Bundesen, 2012).

Bundesen (1990, p. 527) showed that TVA is consistent with the Luce (1963; see also Shepard, 1957)

biased choice rule when assuming that the participant's response is based on the first categorization that completes processing. That is, the first categorization that finishes processing is the one that is preserved in VSTM, while other conflicting categorizations are lost. This solution to mutually exclusive categorization will work well when confusability between possible categorizations is low. A more robust solution is attained through the Poisson Counter Model of identification that we presented in Kyllingsbæk et al. (2012). As explained above, this model assumes that categorizations accumulate independently during the exposure duration and that the categorization with the highest number of counts is the one preserved in VSTM.

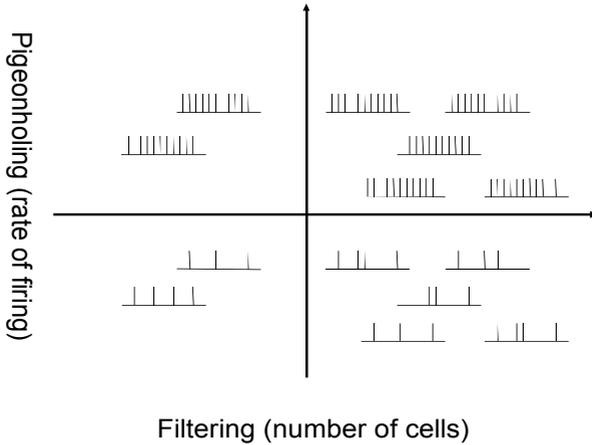


Figure 1: Attentional selection in NTVA: Combined effects of filtering (selection of objects) and pigeonholing (selection of features) on the set of cortical spike trains representing a particular visual categorization of the form ‘object x has feature i ’. The four conditions (quadrants) correspond to the factorial combinations of two levels of filtering (weak vs. strong support to object x) and two levels of pigeonholing (weak vs. strong support to feature i). Filtering changes the number of cortical neurons in which an object is represented. Pigeonholing changes the rate of firing of cortical neurons coding for a particular feature.

3.4 A Neural Interpretation of TVA

In Bundesen et al. (2005, 2011), we proposed a Neural Theory of Visual Attention (NTVA). The aim of the theory is to bridge cognition and neurophysiology by giving a unified mathematical interpretation of attentional behavior at both the level of the organism and single neurons.

A typical neuron in the visual system is assumed to be specialized to represent a single feature and to respond to the properties of only one object at any given time. That is, a neuron can represent different objects at different times, but learning and development aside it always represents the same feature i . Neurons representing feature i are called feature- i neurons. Feature i can be a more or less simple “physical feature” or a “microfeature” in a distributed representation (cf. Hinton, McClelland, & Rumelhart, 1986), and a feature- i neuron may be broadly sensitive to feature i ’s degree of presence rather than sharply tuned to feature i .

The object selection of the neuron occurs by dynamic remapping of the cell’s receptive field so that the functional receptive field contracts around the selected object (cf. Moran & Desimone, 1985; Reynolds, Chelazzi, & Desimone, 1999; Chelazzi, Duncan, Miller, & Desimone, 1998; Chelazzi, Miller, Duncan, & Desimone, 1993). The remapping is done in such a way that the probability that the neuron represents a particular object equals the attentional weight of the object divided by the sum of the attentional weights of all objects in

the classical receptive field (cf. Equations 2 & 3).

Though the total activation representing the categorization depends on the number of neurons representing the categorization, it also depends on the level of activation of the individual neurons representing the categorization. The bias parameter β_i is a scale factor that multiplies activations of all individual feature- i neurons, so the total activation representing the categorization “object x has feature i ” is also directly proportional to β_i (see Martinez-Trujillo & Treue, 2004; McAdams & Maunsell, 1999). By scaling up or down the general processing rate in a population of neurons specifically responsive to certain visual features, the general processing rate of the features for all objects in the visual field will be scaled up or down. Thus, in the neural interpretation of the rate equation (see Equation 2), the total activation representing the categorization “object x has feature i ” is directly proportional to both the number of neurons representing the categorization, and the level of activation of the individual neurons representing the categorization. The number of neurons is controlled by $w_x / \sum w_z$ (filtering), whereas the activation of the individual neurons is controlled by β_i (pigeonholing). The two attentional mechanisms, filtering and pigeonholing, in NTVA are illustrated in Figure 1.

3.5 The Perceptual Cycle in NTVA

In NTVA, the perceptual cycle consists of two waves: *A wave of unselective processing* followed by a *wave of selective processing*. During the first wave, cortical processing is unselective, implying that the processing capacity is distributed at random (unselectively) across the visual field. After the end of the first wave, an attentional weight has been computed for each object in the visual field. The weights are found as levels of activation in a saliency map, which may be located in the pulvinar of the thalamus. The weights are used for redistribution of cortical processing capacity across the objects in the visual field by dynamic remapping of receptive fields, so that the expected number of cells allocated to a particular object becomes proportional to the attentional weight of the object.

During the second wave, cortical processing is selective in the sense that the amount of processing resources (number of neurons) allocated to an object depends on the attentional weight of the object. Because more processing resources are devoted to behaviorally important objects than less important ones, the important objects are more likely to become encoded into VSTM. Furthermore, attentional bias is applied in the second wave of processing, thus scaling up or down firing rates of the cortical neurons, which represent the categorizations of the objects in the visual field. NTVA accounts for a wide range of known attentional effects in human performance (reaction time and error rates) and a wide range

of effects observed in firing rates of single cells in the primate visual system.

3.6 A Theory of Temporal Visual Attention

TVA and NTVA explain attentional behavior at both the level of the organism and at the level of individual visual neurons, but only when the stimuli are presented simultaneously (e.g. whole and partial report). In the natural world however, visual stimuli appear and disappear asynchronously of each other, requiring attention to be distributed not only across space, but also in time. Several different paradigms have been developed to explore allocation of attention in time, e.g. the attentional blink paradigm (Raymond, Shapiro, & Arnell, 1992) and the attentional dwell time paradigm (Duncan, Ward, & Shapiro, 1994). In Petersen et al. (2012), we presented an extension of NTVA—a *Theory of Temporal Visual Attention* (TTVA). When visual stimuli are presented at different points in time, TTVA conjectures that attentional resources are reallocated and a race towards VSTM is initiated each time a new stimulus appears in the visual field. Thus when two target stimuli, T1 and T2, are presented in the attentional dwell time paradigm, two calculations of attentional weights are initiated, one when T1 is presented and another when T2 is presented. After each calculation, the attentional resources are redistributed followed by an attentional race for encoding into VSTM.

During the sequence of presentation, some of the

stimuli may be successfully encoded into VSTM. When this happens, TTVA assumes that the processing resources (cortical visual neurons) that were allocated to process the stimuli are locked temporarily in VSTM. Consequently, the resources are unavailable for redistribution when a new stimulus appears. The locking of resources is the main explanation for the strong impediment in reporting T2 when it follows T1 by less than 500 ms in the attentional dwell time paradigm. The process of locking resources is not instantaneous, but is assumed to be exponentially distributed with a rate parameter λ_l . Resources locked by stimuli encoded in VSTM are released again following a *dwell-time period*, also assumed to be exponentially distributed, but with rate parameter λ_d . When the dwell-time has passed, the neurons locked in VSTM can be redistributed to process new objects, thereby explaining the equivalence in performance on T1 and T2 when presented more than 1000 ms apart. In TTVA, allocation of attentional processing resources and VSTM are highly integrated mechanisms, which cannot be understood in isolation of one another (see further discussion in Sections 5 & 6).

3.7 Application of TVA in Studies of Neuropsychological Patients

During the last ten years, we have extended our effort of modeling behavioral data in normal healthy participants with TVA to modeling of attentional deficits in neuropsychological patients (e.g., Bublak et al., 2005;

Dubois et al., 2010; Duncan et al., 2003; Kyllingsbæk, 2006). Using TVA has several advantages compared to conventional neuropsychological tests of attention for which TVA measures corresponding functions (e.g., Finke et al., 2005). TVA based testing is more sensitive and is therefore able to measure mild attentional deficits that are not evident in standard pen and pencil based neuropsychological tests (e.g., Habekost & Bundesen, 2003). In addition, the test is specific because it gives pure measures of the attentional mechanisms that may be affected. Many neuropsychological tests contain complex motor components, which make it difficult to derive unconfounded measures of the attentional deficits. TVA based testing is also more reliable than standard neuropsychological test measures (Habekost & Starrfelt, 2009). Using bootstrap methods (e.g., Efron & Tibshirani, 1993) it is possible to derive reliability estimates of individual parameters in the TVA model and from these calculate confidence intervals for the parameters of individual patients (see Kyllingsbæk, 2006). Finally, the validity of TVA based testing is high, given the strong theoretical foundation in basic research. TVA has been applied to a wide range of classic findings within experimental psychology of visual attention and identification explaining a wide variety of experimental findings with a very limited set of meaningful parameters (Bundesen, 1990). Quantitative modeling of psychological data, such as modeling used in TVA based testing, is both technically and mathematically challeng-

ing. In Kyllingsbæk (2006), I presented an accessible, user friendly, and flexible program package that enables quantitative fits of TVA to data from whole and partial report experiments. In addition, new computational formulas that were more general than previous ones were presented along with bootstrapping methods to derive estimates of the reliability of the fitted parameters in the model. Recently, Mads Dyrholm developed a more extensive version of the software that can be found at <http://zappa.psy.ku.dk/libtva/> (see also Dyrholm et al., 2011).

4 NTVA and other Theories of Visual Attention

4.1 Feature Integration Theory and the Guided Search Model

During the 80's and early 90's, the *Feature Integration Theory* (FIT) by Treisman and collaborators was very influential within the field of visual attention (e.g., Treisman & Gelade, 1980; Treisman, 1988, 1996). FIT was later followed by the *Guided Search* (GS) model of Wolfe (e.g., Wolfe, 1994; Cave & Wolfe, 1990; Wolfe, Cave, & Franzel, 1989), which is now the dominant theory of visual attention that assumes serial processing. In both FIT and GS, simple stimulus features such as color and orientation are registered automatically, without attention, and in parallel across the visual field. However, registration of objects (items that are defined by conjunctions of features) requires a further stage of processing during which spatial attention is directed to one object at a time. Spatial attention is analog to a spotlight, and the function of directing attention to a particular location in space is to bind all features at the attended location to each other as well as binding them to their location. Thus, localizing a feature requires spatial attention, and only one location can be attended at a time. This contrasts strongly with the basic assumptions in TVA and NTVA of parallel processing at all stages; assumptions which were supported by our

findings in Kyllingsbæk and Bundesen (2007). Here we found evidence that multiple features of two to three stimuli can be identified and localized in parallel. In addition, our findings in Petersen et al. (2012) of attentional dwell times around 500 ms speak against fast serial scanning in visual search assumed by many serial models of visual attention.

4.2 Load Theory

As stated in the beginning of this section, the locus of attentional selection has been discussed intensively since the filter theory was proposed by Broadbent (1958). The *Load Theory* (LT) of Lavie (1995; see also Lavie & Tsal, 1994) was proposed to reconcile this debate. According to LT, perception is capacity-limited and all stimuli are processed in an automatic fashion until this capacity is exhausted (e.g., Lavie, 2005, p. 75). Furthermore, allocation of visual processing capacity happens in two steps: (a) Initial allocation of resources to task-relevant stimuli, followed by (b) automatic allocation (‘spill over’) of the remaining capacity to task-irrelevant stimuli. Two key predictions follow from these premises. First, in situations of high perceptual load (e.g. when many stimuli must be processed), full perceptual capacity will be engaged at the task-relevant stimuli, leaving no spare capacity to process task-irrelevant distractors. Second, in situations of low load (e.g. when few stimuli must be processed and processing is easy), excess capacity not used to process task-relevant stimuli will

automatically ‘spill over’, leading to perception of task-irrelevant distractors. In contrast to LT, both TVA and NTVA assume that attentional resources are allocated in a single step after an attentional weight has been computed for each object in accordance with Equation 3. In Kyllingsbæk, Sy, and Giesbrecht (2011), we presented the results of two experiments that tested the assumptions made by LT and TVA. In both experiments, participants reported targets presented around the perimeter of an imaginary circle centered at fixation while ignoring peripheral distracting flankers. The flankers were presented at two irrelevant locations to the far left and right where no targets appeared. In the first experiment, four targets were shown in every display together with one or two flankers at the irrelevant locations. In addition, we varied the discriminability of the flankers by varying their color (either identical or different from the color of the target letters). In the second experiment, the number of targets and thus the perceptual load was varied systematically. LT predicts that the number and characteristics of flanker stimuli presented at locations known by the observer to always be irrelevant, will not affect performance of the first allocation step in which processing resources are allocated only to stimuli at possible target locations. Contrary to this, TVA predicts only a single allocation step and consequently a systematic effect of both the number and discriminability of the flanker stimuli. TVA accurately fitted the data of the two experiments, thus favoring the simple explanation

with a single step of capacity allocation. In the second experiment, we also showed that the effects of varying perceptual load can only be explained by a combined effect of allocation of processing capacity as well as limits in VSTM as predicted by TVA.

4.3 Biased Competition

TVA bears strong relations to the *Biased Competition* (BC) models of Desimone and Duncan (1995); Desimone (1999); Duncan (1996, 1998); Duncan, Humphreys, and Ward (1997). In fact, TVA may be viewed as one of several different implementations of BC. In BC, attention is thought to operate in parallel across objects in the visual field. In some implementations of BC, activation of a neural unit representing that a given object has a particular feature (e.g., object x is red) facilitates activation of units representing other, unrelated categorizations of the same object (e.g., object x moves upward) and inhibits activation of units representing (related or unrelated) categorizations of objects other than x . This is the *integrated competition hypothesis* of Duncan (1996); units which represent categorizations of the same object reinforce each other while competing against units which represent categorizations of other objects. The formulation suggests that trial-to-trial variations in whether or not a certain feature of a given object is reportable should be (a) positively correlated with whether other features of the same object are reportable and (b) negatively correlated with whether features of other objects

are reportable. This strong version of the biased competition was not supported by our results in Kyllingsbæk and Bundesen (2007). Here we found strong evidence of stochastic independence of encoding of features both within and between several stimuli. Thus, trial-to-trial variations in whether or not a particular feature was correctly reported showed little or no dependence on whether (a) the other feature of the same object was correctly reported or (b) features of the other object were correctly reported.

4.4 The Normalization Model of Attention

In Reynolds and Heeger (2009), the *Normalization Model of Attention* (NMA) was proposed to reconcile alternative theories of attention—theories mainly based on neurophysiological experiments (see also Reynolds et al., 1999). NMA is based on BC and is thus akin to NTVA. With its strong foundation in both BC and neurophysiological data, NMA is particularly interesting to compare with NTVA. In neurophysiological studies of visual attention, the contrast of the stimulus is often systematically manipulated complementary to the systematic variations in exposure duration that have been a major focus in research on TVA/NTVA. Therefore, NMA focuses on effects of attention as modeled by a fixed response gain factor or a change in contrast gain. In addition to this, several neurophysiological studies have implicated an attentional sharpening of neural tuning, which NMA also addresses.

NMA consists of three basic components: A *stimulation field*, an *attention field*, and a *suppressive field*. The stimulation field of a neuron characterizes its selectivity both in terms of spatial position and visual features (e.g. orientation, color, or more complex features). The *stimulation drive* represents the activation of the neuron to the stimuli, varying on these dimensions in the absence of attention and suppression. The definition of the stimulation drive of a neuron is thus analog to the definition of sensory evidence in NTVA, except it is linked to positions rather than objects. The attention field is specified in terms of its gain for each neuron in the population, that is in terms of its spatial and featural extension, and is thus similar to the pertinence values in NTVA. The suppressive field characterizes the spatial locations and visual features that contribute to suppression of the activity in visual neurons. For example, the response of a neuron to a preferred stimulus can be suppressed by the simultaneous presentation of a non-preferred stimulus. The *suppressive drive* is computed from the product of the stimulus drive and the attention field that is then pooled over space and visual features in accordance with the suppressive field. This corresponds to computation of the denominator of the ratio in the rate equation of TVA (see Equation 2). To compute what corresponds to the nominator in the rate equation, the stimulus drive is again multiplied by the attentional field, but without pooling across space and features. Finally, the mechanism of suppression is im-

plemented by dividing (normalizing) the activity given by the product of the stimulation drive and the attention field with the suppressive drive.

The mechanisms of attention in NMA are thus similar in nature to the corresponding mechanisms in NTVA, although the suppressive field provides the NMA with a more flexible mechanism than what is given in the weight ratio in the rate equation of NTVA (but see also Nordfang et al., 2012). There are two other differences between NMA and NTVA that should be emphasized. The first originate in the fact that the attention field is defined in terms of space in NMA, rather than in terms of visual objects, as in NTVA. This enables NMA to explain differences in attentional modulation when the attended stimulus covers the whole receptive field of the neuron recorded in comparison to situations where the stimulus is smaller than the receptive field.

The second is a difference in how NMA and NTVA assume that input is attentionally gated to feature selective neurons—neurons that show effects of attention when several stimuli are presented within its receptive field (e.g., Moran & Desimone, 1985). In NMA, the activity of a neuron is a weighted sum of the input from neurons at earlier stages in the visual system coding for the dominant features of the stimuli within their receptive field (see Reynolds et al., 1999). Thus, the neuron responds simultaneously to input from several visual objects. In contrast, in NTVA the input to the neuron is gated so that the input originates from a single object at

any moment in time. Thus the response of the neuron only codes for features of a single object. This critical difference between NMA and NTVA is due to their respective focus on representation of locations and objects. Mixing information from several objects in the firing rates of cortical neurons as assumed in NMA has a notable weakness. It is not clear how neurons later in the processing chain should use this mixed information from several objects (possibly both targets and distractors) when appropriate actions towards attended objects must be selected (Bundesen et al., 2005, p. 308). The interpretation of firing rates given by NTVA, gives a simpler solution to this problem—a solution where categorizations from attended objects may be forwarded unambiguously to neurons in cortical areas responsible for action selection and execution, while irrelevant information from distractors are efficiently filtered out.

Classic Model of Perception

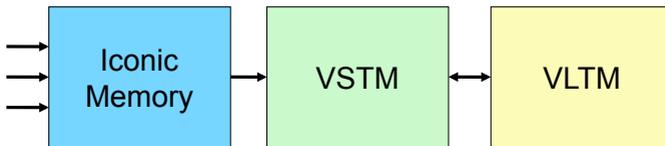


Figure 2: A classic model of visual encoding of information into VSTM and subsequent comparison with VLTM representations.

5 Visual Memory Systems

In many classic models of visual cognition, perception is divided into three distinct stages (see Figure 2). The three stages are: (a) Iconic memory, (b) visual short-term memory (VSTM), and (c) visual long-term memory (VLTM). First, visual information from the eyes is transmitted to iconic memory before it reaches VSTM and VLTM. This temporary store is virtually unlimited in processing and storage capacity, but severely limited in time as it only retains the information for a fraction of a second. Thus, iconic memory represents our immediate sensory impression that vanishes almost instantly when we close our eyes. Iconic memory is the basis for the partial report superiority that Sperling (1960) investigated. The information in iconic memory is then transferred to VSTM through a severely limited processing channel. This channel is directed by attentional processes, such as those indicated in the classic filtering theory by Broadbent (1958). In VSTM, object representations are formed that may then be compared to VLTM representations in order to recognize the identity of the objects. Access to representations in VLTM may only happen through the limited attentional channel and VSTM. Access to these representations happens only one item at time, i.e. in a serial manner (e.g., Treisman & Gelade, 1980).

5.1 Measuring Visual Short-Term Memory Capacity

Traditionally, VSTM capacity has been measured using the whole-report paradigm, in which participants are required to report as many items as possible from a set of briefly presented objects (Cattell, 1885). Sperling (1960, 1963, 1967) used letters as stimuli and found that VSTM capacity was limited to approximately four or five independent items (see also Bundesen, 1990; Bundesen et al., 2005; Kyllingsbæk & Bundesen, 2009; Shibuya & Bundesen, 1988). When familiar, nameable stimuli are used, whole report gives a reliable and easily interpretable measure of VSTM capacity. However, several investigators have objected that the very use of stimuli that can be verbalized may contaminate the estimate of VSTM capacity due to contributions from verbal working memory (Olsson & Poom, 2005; Vogel, Woodman, & Luck, 2001). Moreover, whole report is not feasible for measuring VSTM capacity for unfamiliar stimulus material. For these reasons, the bulk of recent studies of VSTM capacity have used the change detection paradigm (Phillips, 1974). In this paradigm, a memory display containing a varying number of stimuli are presented briefly (< 500 ms) followed by a retention period of 1-3 s where the participant retains the identity and most often the location of the stimuli. Then a probe display is presented containing either all the stimuli from the memory display (Pashler, 1988) or only one of them (Cowan, 2001). In half of the trials one of the

items in the probe display has changed its identity (or the single probe in the paradigm of Cowan, 2001). The task of the participant is to decide whether the probe display is identical to the memory display or a change has occurred. When observing the hits and false alarm rates in the change detection task, an estimate of the VSTM capacity, K , may be obtained. In Kyllingsbæk and Bundesen (2009), we investigated the reliability of the estimated value of K both through repeated experimental measures in the same individual and by mathematical derivation of the predicted variance of the capacity estimates. We first compared the theoretically derived variance to the empirically estimated variance and found close correspondence between predicted and observed variances. Then we showed how the two reliability estimates depended on the number of items to be remembered and the guessing strategy of the observer. The reliability was highest when participants adopted a conservative guessing strategy and lowest when participants guessed ‘change’ and ‘no change’ with equal probabilities (probability matching). Finally, we proposed a new version of the two-alternative choice change detection paradigm, in which the choice is unforced. This new paradigm reduces the variance of the capacity estimate K substantially.

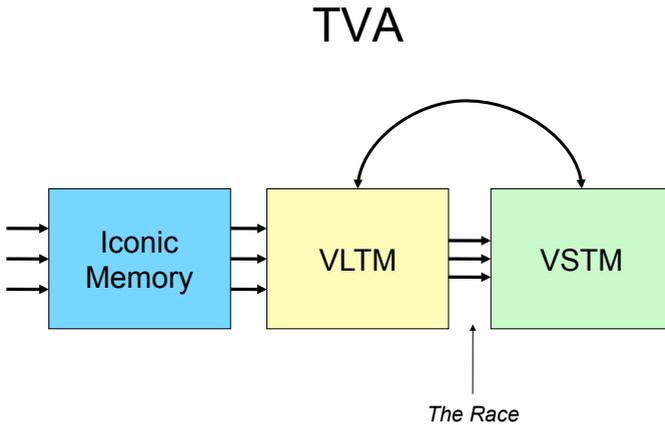


Figure 3: Visual encoding in TVA. The bi-directional arrow between VSTM and VLTM represents comparison processes of items encoded in VSTM and representations in VLTM.

5.2 Visual Short-Term Memory in TVA

Contrary to the classic view of processing in iconic memory, VSTM, and VLTM, the TVA of Bundesen (1990) portrays quite a different view of both the order and the transfer of information between the visual memory systems. This sequence is depicted in Figure 3, where the immediate and most obvious difference is the reversal of the order of processing in VSTM and VLTM (cf. Figure 2). In TVA visual information is first compared to representations in VLTM before selection and encoding into VSTM are made. Furthermore, processing is parallel in

all stages. From iconic memory to VLTM, processing is capacity unlimited, but processing capacity becomes a major limiting factor in the race for storage in VSTM. The core of TVA is this race towards the limited VSTM after the visual impression in iconic memory has been compared to VLTM presentations.

In TVA, VSTM is understood in the classic form as a passive store containing a very limited number of distinct objects. The capacity is not constant over time, but may vary somewhat from moment to moment. Specifically, the capacity is estimated as a probability mixture of two neighboring integer values. Therefore, if the VSTM capacity, K , is estimated at a value of 3.74, it represents the probability mixture that VSTM capacity is three 26% of the time and four 74% of the time. Usually VSTM capacity is approximately four visual objects in normal healthy control participants, but may vary considerably within the normal population, resulting in estimates between two and five objects (see Dyrholm et al., 2011). In Equations 2 and 3 described in the previous section, TVA explains in quantitative details how information is selected into VSTM from where it may influence behavior. More specifically TVA can explain the consequences of both limits in processing capacity and storage capacity simultaneously, thus effectively separating the contribution of these two otherwise entangled effects in the data (cf. Bundesen, 1990; Kyllingsbæk et al., 2011; Shibuya & Bundesen, 1988).

NTVA

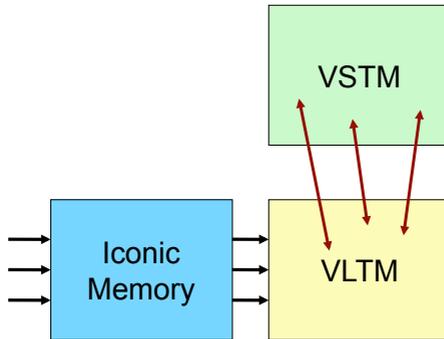


Figure 4: Visual encoding in NTVA.

5.3 Visual Short-Term Memory in NTVA

In the original TVA, VSTM was implemented as a passive store similar to the representation of VSTM in most classic cognitive theories of visual perception and memory. In NTVA however, VSTM is not seen as a place where information is transferred, but rather as a *process* that acts on the information already available in the visual system (see Figure 4). VSTM prolongs the activity in VLTM where it would otherwise have decayed away when the input from the eyes changed or disappeared (see also Cowan, 1995, 2005, for a similar distinction be-

tween VSTM and VLTM). Thus, VSTM is not a passive store of information, but rather a process that enables the visual system to hold on to activity, so that it may associate to other systems in the brain and thus have a global effect leading to coherent overt behavior (e.g., Baars, 2002; Dehaene, Kerszberg, & Changeux, 1998).

The capacity of VSTM in NTVA is also severely limited. However, where the capacity limit in classic models of VSTM is defined in terms of the number of slots, the limit of VSTM in NTVA is in the number of *feedback loops* established between nodes in the store and activated representations of features in VLTM. Thus, following the ideas of Donald Hebb (1949), NTVA implements VSTM as reverberating circuits between nodes in the store and representations in VLTM.

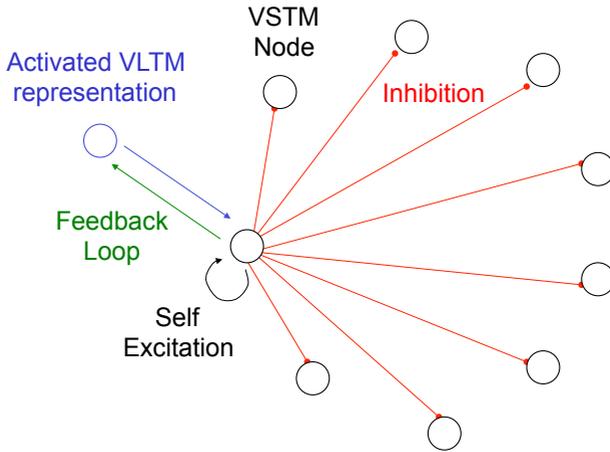


Figure 5: The VSTM network in NTVA. The connections of the VSTM nodes are only shown for a single VSTM node.

As described above, individual variance in VSTM capacity from moment to moment is understood as a probability mixture in TVA. According to NTVA, the variation in VSTM capacity emerges from the structure of the mechanism that implements the process of reverberation. Similar to Usher and Cohen (1999), NTVA suggests that VSTM is implemented as a *winners-take-all* neural network (see Figure 5). In this type of network, each object in the visual field is represented by a node in the network, and each node competes with the other nodes in the network to implement and sustain re-

verberating feedback loops between itself and activated representations in VLTM. Each node in the network has an excitatory connection to itself and inhibitory connection to every other node in the network.

Imagine the situation of whole report, in which a number of letters are presented briefly to a participant who must then report as many of the letters as possible. The letters will activate representations in VLTM that then participate in the attentional processing race in which the representations compete for entrance into VSTM. In NTVA, this means that the categorizations finishing the race will activate nodes in the VSTM network. These nodes will then start establishing feedback loops to the activated representations of the objects in VLTM. At the same time the nodes will increase their own activity by self-excitation as well as inhibit activity of the other nodes in the network. The competition will result in an equilibrium state where only a certain number of nodes are activated, having eliminated the activity in the remaining nodes in the VSTM network.

The number of activated nodes after this *consolidation process* represents the momentary capacity of VSTM. As in the original TVA model, the capacity will vary from moment to moment, but contrary to the probability mixture involving only two neighboring integer values in TVA, the neural network implementation of VSTM in NTVA results in a broader distribution of the capacity. In most cases the capacity of the VSTM will be centered around two neighboring values, but the net-

work implementation of VSTM predicts both lower and higher values for the capacity, although the probability for such events will be smaller (see Dyrholm et al., 2011).

5.4 The Capacity of VSTM—Number of Objects or Amount of Information?

The network implementation of VSTM in NTVA speaks to the heated debate of whether the capacity of VSTM should be measured by the number of objects or the amount of information encoded. According to some researchers, VSTM capacity is restricted by the number of objects encoded whereas the amount of information (e.g. the number of features) encoded for each of these objects is unlimited (e.g., Awh et al., 2007; Luck & Vogel, 1997; Zhang & Luck, 2008). Other researchers have argued strongly that VSTM is not limited by the number of objects encoded, but rather in the total amount of information encoded from the set of objects stored (e.g., Alvarez & Cavanagh, 2004; Bays & Husain, 2008). The network model of VSTM implemented in NTVA may be the mechanism that can accommodate both explanations and thus resolve this controversy. As stated above, the number of items encoded into VSTM is dependent on the dynamics of the nodes competing for activation in the VSTM network. This process is driven by self-excitation and inhibition between the nodes. Following the competition, feedback loops are formed between the active nodes in the VSTM network and the represen-

tations in VLTM that they were initially activated by. This enables activity in the VSTM network and VLTM representations to be sustained once sensory activation of the VLTM representations ceases. However, the reverberating activation of VSTM-VLTM loops is also dependent on the strength and quality of the activation from the VLTM representations that are participating in the initial race for selection, and later encoded as the content in VSTM (i.e. the visual features of the objects). When the stimuli to be remembered are highly similar or when VLTM representations are sparse in number, the feedback loops between these representations and nodes in the VSTM network will be weak. For example, Olsson and Poom (2005) argued that VSTM capacity for novel, highly similar visual shapes and colors was only a single item—significantly below the capacity of VSTM usually found in change detection. Thus the quality and the number of representations in VLTM becomes an indirect but strong determining factor for the capacity of VSTM. Consequently, NTVA predicts that VSTM capacity will be high for familiar well-practiced stimulus materials compared to novel stimuli, stimuli with high complexity, or stimuli that are highly confusable (see Sørensen & Kyllingsbæk, 2012). This may explain why the capacity for distinct primary colors or well-known letters from the Roman alphabet are larger than the capacity for unknown Chinese characters (in participants from a Western population) or highly similar 3D shapes (Alvarez & Cavanagh, 2004).

NTVA

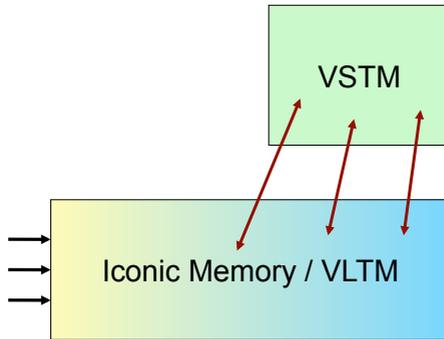


Figure 6: Iconic memory as activated representations in VLTM.

5.5 Iconic Memory as Activated VLTM Representations

Finally, a further development of the architecture of the visual memory system is proposed in Figure 6. In both classic models of visual perception as well as in TVA and NTVA, visual memory is divided into three distinct stores with different properties. In Figure 6, a different interpretation is suggested: Iconic memory and VLTM are not implemented as distinct structural stores. Rather, iconic memory and VLTM are two aspects of the same memory system. Specifically, iconic memory

is the activated subset of representations in VLTM—the structure where processing is taking place. In terms of the rate and weight equations (see Equations 2 & 3) of TVA/NTVA, the total set of $\eta(x, i)$ values represents the content of iconic memory. In this interpretation, VLTM should be understood in broad terms. This includes very simple visual features, such as representations of orientation and spatial frequency in the primary visual cortex, as well as high-level visual representations of complex forms in the inferotemporal cortex. If iconic memory is our visual impression of the outside world, the conjecture is that we see the world through the representations in our VLTM. Some of these representations are very simple and implemented genetically or formed very early in life, yet at higher levels the representations are mainly formed by visual experiences and may be influenced, changed, or developed anew over relatively short periods of time (e.g., Kobatake, Wang, & Tanaka, 1998).

6 An Integrated View of Visual Cognition

The previous sections outlined components of a general theoretical framework for understanding visual encoding, attention, and short-term memory. In this framework, the encoding of visual information happens independently and in parallel across both features and objects in the visual field (Kyllingsbæk & Bundesen, 2007). Allocation of visual processing resources is understood in terms of dynamic remapping of the receptive fields of cortical neurons tuned to different visual categories/features (Bundesen et al., 2005). When several stimuli are presented within the same receptive field, competition for processing resources will be strong. On other hand, when only a single stimulus is presented within each receptive field, allocation of processing resources will be independent of the number of stimuli that are being processed (Kyllingsbæk et al., 2007). Encoding and identification of a stimulus is a cumulative Poisson process in which evidence for different possible categorizations of the stimulus is collected while the stimulus is physically available to the visual system. Evidence for each of the categories will be aggregated in a counter and the categorization accumulating the highest number of counts will be selected for the final choice of categorization (Kyllingsbæk et al., 2012).

When encoding information from a complex visual scene, attending to the important stimuli and ignoring

irrelevant distractors is vital. Thus the outlines of basic encoding principles must be embraced in a broader theoretical framework for understanding visual processing when both relevant and irrelevant information is present. In NTVA, we developed a computational theory of visual attention whereby processing resources are represented as the total firing rate in the population of cortical neurons in the visual system (Bundesen et al., 2005). These processing resources may be allocated by dynamic remapping of the cortical neurons and scaled up or down by multiplicative scaling of the firing rates via adjustment of perceptual bias for different categorizations. The three characteristics of the encoding process described above follow naturally from the basic assumptions in NTVA.

According to NTVA, visual information is encoded into a VSTM store. However, in contrast to most classic theories of visual attention, VSTM in NTVA is not a static storage mechanism where the result of the encoding processes is retained temporarily. In NTVA, VSTM is implemented as a winners-take-all network. A competitive process implemented by self-excitation and mutual inhibition leads to equilibrium states in which a small number of visual objects are represented in the store. In addition, the VSTM network consists only of feedback loops that sustain activated VSTM representations by reverberation—representations that were activated by the encoding process. Thus, the content of VSTM is not stored in the VSTM network as such, but

as activated representations in VLTM.

In combination with the theoretical interpretations of encoding and attention, realizing VSTM as a process in interaction with representations in VLTM rather than a passive storage mechanism, offers a fundamental change in our understanding of how visual information is processed. Visual encoding, attention, and short-term memory are not isolated stages in a processing chain where information is handed over from one mechanism to the next in sequential order. Rather, the present theoretical position conjectures that encoding, attention and short-term memory are all processes working in concert on representations in VLTM by activation, attentional remapping of receptive fields, feature specific biasing of cortical populations of neurons, and reverberation to sustain activated representations when the stimulus changes or is not physically present any more.

7 Danish Summary

Emnet for nærværende afhandling er visuel kognition generelt og specifikt udviklingen af matematiske modeller for visuel perception, opmærksomhed og korttidshukommelse. De matematiske modeller blev konstrueret med den hensigt, at de skulle være så simple som mulige og stadig have en høj grad af validitet givet ved deres struktur og parametre. Sidstnævnte bestemmes af forbindelsen mellem parametrene og psykologisk meningsfulde begreber, som fx indkodningshastighed og hukommelseskapacitet. Ydermere var formålet med min forskning at udvide modellerne, således at de kan forklare de neurale mekanismer, som danner det biologiske grundlag for de nævnte kognitive processer.

Den første del af afhandlingen giver en præsentation af eksperimentelle og teoretisk studier vedrørende indkodning af visuel information, når en eller flere stimuli identificeres. Dette efterfølges af en gennemgang af en neural teori for visuel opmærksomhed. Endelig diskuteres et integreret syn på visuel hukommelse, hvor forbindelserne mellem sensorisk hukommelse, visuel korttidshukommelse og visuel langtidshukommelse nytænkes.

8 References

- Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, *15*, 106–111.
- Awh, E., Barton, B., & Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science*, *18*, 622–628.
- Baars, B. J. (2002). The conscious access hypothesis: origins and recent evidence. *Trends in Cognitive Sciences*, *6*, 47–52.
- Bartels, A., & Zeki, S. (2004). Functional brain mapping during free viewing of natural scenes. *Human Brain Mapping*, *21*, 75–85.
- Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science*, *321*, 851–854.
- Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature*, *226*, 177–178.
- Broadbent, D. E. (1958). *Perception and communication*. London, UK: Pergamon press.
- Broadbent, D. E. (1971). *Decision and stress*. London, UK: Academic Press.

Bublak, P., Finke, K., Krummenacher, J., Preger, R., Kyllingsbæk, S., Müller, H. J., et al. (2005). Usability of a theory of visual attention (tva) for parameter-based measurement of attention ii: Evidence from two patients with frontal or parietal damage. *Journal of the International Neuropsychological Society*, *11*, 843–854.

Bundesen, C. (1990). A theory of visual attention. *Psychological Review*, *97*, 523–547.

Bundesen, C., Habekost, T., & Kyllingsbæk, S. (2005). A neural theory of visual attention: bridging cognition and neurophysiology. *Psychological Review*, *112*, 291–328.

Bundesen, C., Habekost, T., & Kyllingsbæk, S. (2011). A neural theory of visual attention and short-term memory (ntva). *Neuropsychologia*, *49*, 1446–1457.

Bundesen, C., & Harms, L. (1999). Single-letter recognition as a function of exposure duration. *Psychological Research*, *62*, 275–279.

Bundesen, C., Kyllingsbæk, S., Houmann, K. J., & Jensen, R. M. (1997). Is visual attention automatically attracted by ones own name? *Perception & Psychophysics*, *59*, 714–720.

Cattell, J. M. (1885). Über die zeit der erkenntung und benennung schriftzeichen, bildern und farben. *Philosophische Studien*, *2*, 635–650.

Cave, K. R., & Wolfe, J. M. (1990). Modeling the role of parallel processing in visual search. *Cognitive Psychology*, *22*, 225–271.

Chelazzi, L., Duncan, J., Miller, E. K., & Desimone, R. (1998). Responses of neurons in inferior temporal cortex during memory-guided visual search. *Journal of Neurophysiology*, *80*, 2918–2940.

Chelazzi, L., Miller, E. K., Duncan, J., & Desimone, R. (1993). A neural basis for visual search in inferior temporal cortex. *Nature*, *363*, 345–347.

Cowan, N. (1995). *Attention and memory: An integrated framework*. New York, NY: Oxford University Press.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*, 87–185.

Cowan, N. (2005). *Working memory capacity*. New York, NY: Psychology Press.

Dehaene, S., Kerszberg, M., & Changeux, J.-P. (1998). A neuronal model of a global workspace in effortful cognitive tasks. *Proceedings of the National Academy of Sciences*, *95*, 14529–14534.

Desimone, R. (1999). Visual attention mediated by biased competition in extrastriate cortex. In G. W. Humphreys, J. Duncan, & A. Treisman (Eds.),

Attention, space, and action (pp. 13–31). Oxford, UK: Oxford University Press.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*, 193–222.

Desimone, R., & Gross, C. G. (1979). Visual areas in the temporal cortex of the macaque. *Brain Research*, *178*, 363–380.

Desimone, R., & Ungerleider, L. G. (1989). Neural mechanisms of visual perception in monkeys. In F. Boller & J. Grafman (Eds.), *Handbook of neuropsychology* (Vol. 2, pp. 267–299). Amsterdam, Netherlands: Elsevier.

Deutsch, J. A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review*, *70*, 80–90.

Dubois, M., Kyllingsbæk, S., Prado, C., Musca, S. C., Peiffer, E., Lassus-Sangosse, D., et al. (2010). Fractionating the multi-character processing deficit in developmental dyslexia: Evidence from two case studies. *Cortex*, *46*, 717–738.

Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General*, *113*, 501–517.

Duncan, J. (1996). Cooperating brain systems in selective perception and action. In T. Inui & J. L. McClelland (Eds.), *Attention and performance xvi: Information integration in perception and communication* (pp. 549–578). Cambridge, MA: MIT Press.

Duncan, J. (1998). Converging levels of analysis in the cognitive neuroscience of visual attention. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *353*, 1307–1317.

Duncan, J., Bundesen, C., Olson, A., Humphreys, G., Ward, R., Kyllingsbæk, S., et al. (2003). Attentional functions in dorsal and ventral simultanagnosia. *Cognitive Neuropsychology*, *20*, 675–701.

Duncan, J., Humphreys, G., & Ward, R. (1997). Competitive brain activity in visual attention. *Current Opinion in Neurobiology*, *7*, 255–261.

Duncan, J., Ward, R., & Shapiro, K. (1994). Direct measurement of attentional dwell time in human vision. *Nature*, *369*, 313–315.

Dyrholm, M., Kyllingsbæk, S., Espeseth, T., & Bundesen, C. (2011). Generalizing parametric models by introducing trial-by-trial parameter variability: The case of tva. *Journal of Mathematical Psychology*, *55*, 416–429.

Efron, B., & Tibshirani, R. (1993). *An introduction to the bootstrap* (Vol. 57). New York, NY: Chapman & Hall/CRC.

Finke, K., Bublak, P., Krummenacher, J., Kyllingsbæk, S., Müller, H. J., & Schneider, W. X. (2005). Usability of a theory of visual attention (tva) for parameter-based measurement of attention i: Evidence from normal subjects. *Journal of the International Neuropsychological Society*, *11*, 832–842.

Habekost, T., & Bundesen, C. (2003). Patient assessment based on a theory of visual attention (tva): Subtle deficits after a right frontal-subcortical lesion. *Neuropsychologia*, *41*, 1171–1188.

Habekost, T., & Starrfelt, R. (2009). Visual attention capacity: A review of tva-based patient studies. *Scandinavian Journal of Psychology*, *50*, 23–32.

Hebb, D. O. (1949). *The organization of behavior: A neuropsychological theory*. New York, NY: Wiley.

Hinton, G. E., McClelland, J. L., & Rumelhart, D. E. (1986). Distributed representations. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing* (Vol. 1, pp. 77–109). Cambridge, MA: MIT Press.

Kobatake, E., Wang, G., & Tanaka, K. (1998). Effects of shape-discrimination training on the selectiv-

ity of inferotemporal cells in adult monkeys. *Journal of Neurophysiology*, *80*, 324–330.

Kyllingsbæk, S. (2001). *Parallel and serial selective processing in vision*. Unpublished doctoral dissertation, University of Copenhagen, Copenhagen, Denmark.

Kyllingsbæk, S. (2006). Modeling visual attention. *Behavior Research Methods*, *38*, 123–133.

Kyllingsbæk, S., & Bundesen, C. (2007). Parallel processing in a multifeature whole-report paradigm. *Journal of Experimental Psychology: Human Perception and Performance*, *33*, 64–82.

Kyllingsbæk, S., & Bundesen, C. (2009). Changing change detection: Improving the reliability of measures of visual short-term memory capacity. *Psychonomic Bulletin & Review*, *16*, 1000–1010.

Kyllingsbæk, S., Markussen, B., & Bundesen, C. (2012). Testing a poisson counter model for visual identification of briefly presented, mutually confusable single stimuli in pure accuracy tasks. *Journal of Experimental Psychology: Human Perception and Performance*, *38*, 628–642.

Kyllingsbæk, S., Sy, J. L., & Giesbrecht, B. (2011). Understanding the allocation of attention when faced with varying perceptual load in partial report: A computational approach. *Neuropsychologia*, *49*, 1487–1497.

- Kyllingsbæk, S., Valla, C., Vanrie, J., & Bundesen, C. (2007). Effects of spatial separation between stimuli in whole report from brief visual displays. *Perception & Psychophysics*, *69*, 1040–1050.
- Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 451–468.
- Lavie, N. (2005). Distracted and confused?: Selective attention under load. *Trends in Cognitive Sciences*, *9*, 75–82.
- Lavie, N., & Tsal, Y. (1994). Perceptual load as a major determinant of the locus of selection in visual attention. *Perception & Psychophysics*, *56*, 183–197.
- Logan, G. D. (1996). The code theory of visual attention: An integration of space-based and object-based attention. *Psychological Review*, *103*, 603–649.
- Luce, R. D. (1963). Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (pp. 103–189). New York, NY: Wiley.
- Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279–281.
- Marr, D. (1982). *Vision*. San Francisco, CA: W.H. Freeman and Company.

- Martinez-Trujillo, J. C., & Treue, S. (2004). Feature-based attention increases the selectivity of population responses in primate visual cortex. *Current Biology*, *14*, 744–751.
- McAdams, C. J., & Maunsell, J. H. (1999). Effects of attention on orientation-tuning functions of single neurons in macaque cortical area v4. *The Journal of Neuroscience*, *19*, 431–441.
- Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science*, *229*, 782–784.
- Nordfang, M., Dyrholm, M., & Bundesen, C. (2012). Identifying bottom-up and top-down components of attentional weight by experimental analysis and computational modeling. *Journal of Experimental Psychology: General*.
- Olsson, H., & Poom, L. (2005). Visual memory needs categories. *Proceedings of the National Academy of Sciences of the United States of America*, *102*, 8776–8780.
- Pashler, H. (1988). Familiarity and visual change detection. *Perception & Psychophysics*, *44*, 369–378.
- 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 perception and performance*, *38*, 498–514.

- Petersen, A., Kyllingsbæk, S., & Bundesen, C. (2012). Measuring and modeling attentional dwell time. *Psychonomic Bulletin & Review*, *19*, 1029–1046.
- Phillips, W. (1974). On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics*, *16*, 283–290.
- Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 849–860.
- Reynolds, J. H., Chelazzi, L., & Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4. *The Journal of Neuroscience*, *19*, 1736–1753.
- Reynolds, J. H., & Heeger, D. J. (2009). The normalization model of attention. *Neuron*, *61*, 168–185.
- Shepard, R. N. (1957). Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika*, *22*, 325–345.
- Shepard, R. N. (1958a). Stimulus and response generalization: Deduction of the generalization gradient from a trace model. *Psychological Review*, *65*, 242–256.

Shepard, R. N. (1958b). Stimulus and response generalization: Tests of a model relating generalization to distance in psychological space. *Journal of Experimental Psychology*, *55*, 509–523.

Shibuya, H., & Bundesen, C. (1988). Visual selection from multielement displays: measuring and modeling effects of exposure duration. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 591–600.

Sørensen, T. A., & Kyllingsbæk, S. (2012). Short-term storage capacity for visual objects depends on expertise. *Acta Psychologica*, *140*, 158–163.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied*, *74*, 1–29.

Sperling, G. (1963). A model for visual memory tasks. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *5*, 19–31.

Sperling, G. (1967). Successive approximations to a model for short term memory. *Acta Psychologica*, *27*, 285–292.

Treisman, A. (1988). Features and objects: The fourteenth bartlett memorial lecture. *The Quarterly Journal of Experimental Psychology*, *40*, 201–237.

Treisman, A. (1996). The binding problem. *Current Opinion in Neurobiology*, *6*, 171–178.

- Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*, 97–136.
- Usher, M., & Cohen, J. D. (1999). Short term memory and selection processes in a frontal-lobe model. In D. Heinke, G. W. Humphreys, & A. Olson (Eds.), *Connectionist models in cognitive neuroscience* (pp. 78–91). Berlin, Germany: Springer Verlag.
- Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology Human Perception and Performance*, *27*, 92–114.
- Wandell, B. A. (1995). *Foundations of vision*. Sunderland, MA: Sinauer Associates.
- Wheeler, M., & Treisman, A. (2002). Binding in visual short term memory. *Journal of Experimental Psychology: General*, *131*, 48–64.
- Wolfe, J. M. (1994). Guided search 2.0 a revised model of visual search. *Psychonomic Bulletin & Review*, *1*, 202–238.
- Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human perception and Performance*, *15*, 419–433.

Zeki, S. (1993). *A vision of the brain*. Oxford, UK: Blackwell Scientific Publications.

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*, 233–235.