Department of Psychology

University College London

University of London

How many can you see at a glance?

The Role of Attention in Visual Enumeration

Petra Vetter

Supervisors

Prof. Brian Butterworth & Prof. Alan Johnston

Submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy
University College London

June 2009

I, Petra Vetter confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

There is considerable controversy as to how the brain extracts numerosity information from a visual scene and as to how much attention is needed for this process. Traditionally, it has been assumed that visual enumeration is subserved by two functionally distinct mechanisms: the fast and accurate apprehension of 1 to about 4 items, a process termed "subitizing", and the slow and error-prone enumeration of larger numerosities referred to as "counting". Further to a functional dichotomy between these two mechanisms, an attentional dichotomy has been proposed. Subitizing has been thought of as a pre-attentive and parallel process, whereas counting is supposed to require serial attention.

In this work, the hypothesis of a parallel and pre-attentive subitizing mechanism was tested. To this aim, the amount of attention that could be allocated to an enumeration task was experimentally manipulated. In Experiment 1, attentional set was manipulated such that attention could either be drawn to the relevant of two subsets to enumerate or had to be distributed to both subsets. Furthermore, the relationship of enumeration to perceptual grouping and item discrimination was explored. In Experiment 2, a dual-task approach was employed in which the amount of attentional resources available to enumeration was systematically modulated by imposing an additional task and by varying its attentional load. Experiment 3 investigated the neural correlates of visual enumeration under attentional load using functional magnetic resonance imaging (fMRI).

Results indicated that (1) enumeration, particularly subitizing, was clearly compromised under conditions of distributed or reduced attention. (2) Both the enumeration of small and large numerosities was affected by such attentional manipulations. (3) Subitizing selectively activated brain areas associated with stimulus-driven attention. (4) Enumeration is contingent on other potentially attention-demanding visual processes such as perceptual grouping. The evidence presented here seriously challenges the traditionally held claim of a parallel and pre-attentive subitizing mechanism and suggests instead that small numerosity judgement requires visual attention. This weakens the argument of an attentional as well as a functional dichotomy and strengthens the idea that enumeration may be subserved by a single, continuous mechanism.

Acknowledgements

My time at the Institute of Cognitive Neuroscience at UCL has been a very exceptional one. I feel privileged to have been able to work with so many extraordinary bright, creative, passionate, open-minded and most importantly, warm and interesting people. My time at the ICN has been both an intellectually challenging and inspiring as well as a personally very enriching experience.

I wish to thank Brian Butterworth for his warm welcome in his group, for giving me the opportunity to work at the ICN and be part of the European research training network NUMBRA. I am very grateful for his loyalty and support, and particularly for the scientific freedom he gave me. He gave me the freedom to independently choose the research question for my PhD, to realise my scientific ideas, to develop my own experimental approaches and to collaborate with many different people within and outside the ICN. This is particularly remarkable as the current project challenges an aspect of his own theory. His open-minded attitude gave me an excellent example for scientific debate and strengthened my intellectual independence considerably.

Alan Johnston has given many thoughtful comments and good advice, I also wish to thank him for his kindness and constant support.

I am deeply indebted to Bahador Bahrami. He has spent innumerable hours and infinite patience to answer all the many questions I had, to teach me programming, psychophysics, fMRI design and SPM and to share his extensive theoretical knowledge about visual perception. Much of the work presented in this thesis has been developed together with him (especially Experiments 2 and 3). I particularly enjoyed our "fights" (or rather, passionate scientific discussions) and am very grateful for having had such a sharp and skillfully arguing counterpart to sharpen my own scientific argumentation. Apart from his brilliant mind, his warm, welcoming and kind manner made him a good friend and our collaboration a true pleasure.

Ashish Ranpura has helped me with programming, graphical issues and language advice and in particular, has been a truly wonderful and supporting friend during the whole time of my PhD. Marinella Cappelletti has provided a lot of support on all kinds of issues, scientific

4

Contents

List of Figures

List of Tables

Chapter 1

General Introduction

1.1   Is there a "Magical Number 4"?

"Supposing that the mind is not limited to the simultaneous consideration of a single object, a question arises: How many objects can it embrace at once?"

Sir William Hamilton, 1866

In 1871, Stanley Jevons was the first who addressed Hamilton's question in a self-administered psychological experiment: he threw a variable number of black beans into a white card box and as soon as they came to rest, he rapidly estimated their number "without the least hesitation". He underwent over 1000 trials and discovered that he made no errors when there were only 3 or 4 beans in the box, but that his judgement became increasingly error prone with 5 beans or more. He concluded that "the number five is beyond the limit of perfect discrimination" (Jevons, 1871).

The data pattern observed today with the aid of modern technology is in principle the same as in Jevon's study: subjects are usually very fast and precise to enumerate up to 4 items and become increasingly slow and inaccurate when asked to estimate the number of 5 or more items (see Fig. 1.1 for an example of reaction time and error rate patterns as a function of the number of items to enumerate). However, no satisfying explanation for this behavioural
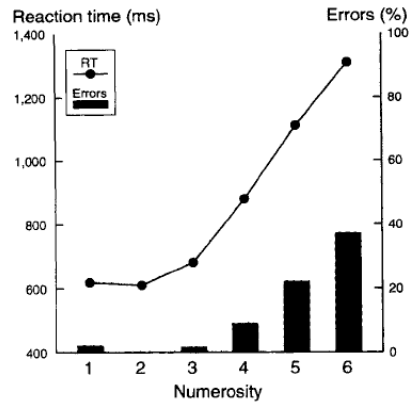
Figure 1.1: Typical Enumeration Performance

Typical pattern of reaction times and error rates as a function of the number of items
to enumerate. Taken from Dehaene & Cohen (1994).

phenomenon has been provided yet and it is still controversially debated why it is so easy to
enumerate up to 4 items, and what makes the estimation of more than 5 items so difficult
for the brain. Is there something special, or even "magical" about the number 4 as Cowan
(2001) suggests? What are the brain mechanisms responsible for the extraction of numerical
information in a visual percept and are they different for small and large numerosities?

Intuitively, it appears that small numerosities can be perceived "at a glance" without
the laborious one-by-one counting that is needed for higher numerosities. Many scientists have
suggested that visual enumeration is divided into two functionally specialised mechanism: one
for numerosities up to 3 or 4 items and one for more than 5 items. Kaufman et al. (1949) coined
a specific term for the fast apprehension of small numerosities: "subitizing". The term is derived

from latin *subitus*, meaning "sudden", reflecting the perception at "a glance". For the enumeration of numerosities above 4, the terms "estimating" or "counting" are used. "Estimation" refers to the error-prone but fast approximation of numerosity, whereas "counting" reflects a learned verbal mechanism where each item receives a label in one-to-one correspondence (Kaufman et al., 1949).

The special character of the subitizing process has been emphasised by many scientists. Bourdon (1908) suggested that small numerosities are apprehended as effortlessly as colour or shape, the "twoness" of a group of items is perceived in the same way as its "redness", for example. Butterworth (1999) proposes that subitizing is an important component of a hard-wired, innate "number module" with which the human brain is endowed. Also Feigenson, Dehaene and Spelke (2004) suggest a "core system" specialised for the precise representation of small numerosities that is distinct from a second system for large and approximate number representations. The main arguments for these latter theories come from finding of a set-size limit in infants and animals, but also from evidence in patients with brain damage and healthy adults as reviewed below.

The idea of two functionally specialised enumeration mechanisms has led to the common assumption that the observed reaction time and error rate patterns are best fitted with two discontinuous linear functions, a rather flat one for small numerosities and a steeper one for larger numerosities. Further to a functional dichotomy between subitizing and counting, the bilinear fit of performance slopes has led to the conclusion that both processes are separated also by an attentional dichotomy. On the basis of the interpretation in classical visual search studies (e.g. Treisman & Gelade, 1980), flat reaction time slopes have been taken as evidence for parallel and pre-attentive processing whereas sloped performance curves are thought to reflect serial and attentive processing. By analogy, the rather flat reaction time slope for subitizing has led to the conclusion that it is a parallel and pre-attentive process (Sagi & Julesz, 1985), whereas the steeper counting slope was taken as indication for a serial and attentive process. The attentional dichotomy is conceptually different from the functional one (one does not necessarily imply the other), but as they emerged from the same assumption (the bilinear fit of performance curves), they are tightly linked.

However, other researchers have argued that the assumption of two distinct enumeration mechanisms for small and large numerosities is a conceptual artefact and have instead argued for a single mechanism. In the following, the arguments for and against both a functional as well as an attentional dichotomy are reviewed, along with the attempts to explain the characteristic data pattern and the presumed special status of subitizing. This thesis focuses on the attentional requirements of the subitizing process, thus particular emphasis is given on this aspect of the literature.

1.2 Arguments for a Subitizing - Counting Dichotomy

1.2.1 Behavioural Studies

Many studies have replicated the typical data pattern of fast and accurate enumeration of small numerosities and increasingly slow and error-prone enumeration of larger numerosities under different experimental conditions (e.g. Taves, 1941; Kaufman et al., 1949; Jensen, Reese & Reese, 1950; Atkinson, Campbell & Francis, 1976; Mandler & Shebo, 1985; Trick & Pylyshyn, 1993; Simon & Vaishnavi, 1996). Typically, reaction time slopes of about 40-100 ms per item are found for subitizing and slopes of about 250-350 ms/item are described for counting (see Trick & Pylyshyn, 1994, for a review). The discontinuity has also been observed in error rates/accuracy and confidence ratings (Kaufman et al., 1949; Taves, 1941).

Taves (1941) was the first to apply a bilinear fit to the data pattern and to suggest that two separable process may underly the perception of "numerousness". Kaufman et al. (1949) as well as Jensen, Reese and Reese (1950) followed his interpretation, though in these early studies, the set size limit was still thought to be at 6 items.

Atkinson, Campbell and Francis (1976) found clear evidence for a capacity limit at 4 items at relatively short stimulus presentation times. However, they were the first to discover that the subitizing limit can be smaller depending on the stimulus properties. When decreasing the spacing between linearly arranged elements, they discovered that the subitizing limit can fall to 2 items. Mandler and Shebo (1982) replicated the typical data pattern and suggest the

subitizing limit to be at 3 items.

It is nowadays agreed that the subitizing limit lies at 3 or 4 items rather than 6; however the question of whether it is exactly at 3 or 4 items is a rather fruitless discussion. The data in almost all studies show that the kink in the performance slope is not a very sharp one and it is well established that there are individual differences in the subitizing limit between subjects (e.g. Akin & Chase, 1978; Trick & Pylyshyn, 1994). When averaging across subjects, it is unavoidable that sharp discontinuities are smoothed. Furthermore, the discontinuity depends on experimental conditions and stimulus properties (see Atkinson, Campbell & Francis, 1976) as well as on the bilinear data fit that researchers apply. Given that the majority of studies define the subitizing range from 1 to 4 (see Trick & Pylyshyn, 1994 and references within), the same range will be assumed here.

Sagi and Julesz (1985) were the first to use subitizing in a visual search paradigm and to interpret the shallow reation time slope as an indicator of a parallel and pre-attentive process. Within a display of differently orientated lines, they found that detecting up to four targets amongst distractors and deriving their numerosity yielded flat reaction time curves whereas orientation discrimination of the same targets yielded sloped curves. In line with the attention literature at the time (e.g. Treisman & Gelade, 1980; Julesz, 1984), the authors concluded that numerosity judgement is performed pre-attentively whereas orientation judgement requires serial attention.

Both Simon and Vaishnavi (1996) and Atkinson, Campbell and Francis (1976) studied enumeration performance in afterimages and thus under conditions of disabled eye-movements. They replicated the typical performance pattern: error-free enumeration up to 4 dots and error-prone enumeration above 4 dots. Simon and Vaishnavi (1996) drew specific conclusion from these results with respect to the attentional requirements of enumeration: they inferred that enumerating large numerosities perfectly accurately requires the eyes, and thus the attentional focus, to be shifted through the display. On this basis, they argue for differential attentional mechanisms in both enumeration modes.

1.2.2   Neuropsychological Studies

A double dissociation between subitizing and counting would be a strong argument for a functional dichotomy; i.e. if there were brain-damaged patients unable to subitize and able to count and patients with different lesions able to subitize and unable to count.

Butterworth (1999) described a patient suffering from developmental dyscalculia, who counted even small number of items (at a rate of about 200 ms/item) when presented with dot arrays for an unlimited amount of time. It appeared that he might be unable to subitize. However, when presented with the stimulus very briefly (which forced him to estimate), he showed the same discontinuity in his reaction time slope at around 4 items as normals (Piazza et al., 2002). Piazza and Butterworth concluded that his behaviour in unlimited displays was due to a lack in confidence to estimate items, given his general problems in numerical tasks. Another case of aquired acalculia was also unable to give an immediate estimate of small numerosities unless allowed to use finger pointing and verbal counting (Cipolotti et al., 1991). However, this patient could not recall numbers above 4 and her ability to count higher numerosities could not be assessed. Therefore, a discontinuity in performance between small and large numbers (or a lack of such) could not be determined in this case. Thus, the neuropsychological evidence for an inability to subitize and an ability to count remains unclear.

Dehaene and Cohen (1994) provided evidence for the opposite dissociation: ability to subitize and inability to count. They also argue for dissociable attentional mechanisms for subitizing and counting. They investigated enumeration and visual search performance in patients with simultanagnosia (the inability to recognize multiple elements in a simultaneously displayed visual presentation). These patients showed preserved subitizing ability for up to three items but failed to enumerate larger sets. The authors concluded that the patients were unable to count large sets because they were unable to keep track of the spatial locations of counted items resulting in incorrect running totals. As the patients were comparably accurate in enumerating small sets, the authors argue that subitizing must be independent of serial attention, and relies on a spatially parallel mechanism. Dehaene later adopted this view of pre-attentive subitizing in his theoretical framework on number cognition (Dehaene, 1997).

Vuilleumier and Rafal (2000) claim to have found neuropsychological evidence specifically for a pre-attentive subitizing mechanism when testing patients with hemispatial neglect on enumeration, localisation and object discrimination. Patients with hemispatial neglect typically have lesions in right parietal cortex and are unable to attend to the left (i.e. contralesional) side of space. Theses patients showed marked contralesional extinction (i.e. they performed at chance) when asked to report the location or the presence of certain stimuli presented to one or both visual hemifields. However, enumeration of up to four items caused no significant extinction when items were distributed across both hemifields. Reaction time for enumeration was similar on trials when items were presented in the intact field or in both the intact and the impaired field. These results suggest that patients were able to derive the numerosity of a set although they were not consciously aware of all the individual items and their location. Vuilleumier and Rafal (2000) regarded this finding as consistent with a parallel subitizing process where a small set of stimuli is individuated and grouped without spatial attention and thus survives extinction.

1.2.3   Animal and Infant Studies

As mentioned above, Feigenson, Dehaene and Spelke (2004) argue for a specialised "core system" for small numerosities based on evidence from infant and animal studies. One of the arguments that speaks for their theory, for example, is that 5 month-olds can discriminate 2 versus 3 dots in a habituation task, but not larger collections of the same ratio (i.e. 4 versus 6 dots; Starkey & Cooper, 1980). Also untrained rhesus monkeys distinguish between 1, 2 and 3 items but fail to discriminate between 4 versus 5 or even 3 versus 8 (Hauser, Carey & Hauser, 2000).

Nieder and colleagues investigated visual enumeration in rhesus monkeys more in detail, both behaviourally as well as neurophysiologically (Nieder, Freedman & Miller, 2002; Nieder & Miller, 2004). In both studies, monkeys were trained to discriminate numerosities from 1 up to 5 (Nieder, Freedman & Miller, 2002) or 1 to 7 (Nieder & Miller, 2004). Although Nieder and colleagues suggest a continuum between small and large numerosities rather than a subitizing-counting dichotomy, they conclude that enumerating small sets must occur in parallel. One of their arguments is that numerosity-tuned neurons in the monkey pre-frontal cortex respond with a constant latency after stimulus onset throughout all numerosities 1 to 5 (Nieder, Freedman &

Miller, 2002).[1] Furthermore, the monkeys' number of eye-movements do not increase with set size (Nieder & Miller, 2004). However, it has to be pointed out that the monkeys' performance with respect to reaction time and eye movements was at chance for numerosities above four. Thus, the results do not allow conclusions about the processing of larger numerosities and whether it is qualitatively different from smaller numerosities.

1.2.4   Imaging Studies

Sathian et al. (1999) were the first to conduct a PET (positron emission tomography) study on the issue and claim to have found evidence for both a functional as well as an attentional dichotomy on a neural level. Subitizing resulted in neural activation of a few areas in occipital extrastriate cortex, whereas counting activated a much more widespread, higher-level network in frontal and parietal areas. The authors interpreted these results as evidence for distinguishable neural systems in enumeration, with subitizing only recruiting areas associated with early pre-attentive processes in visual cortex and counting activating areas involved in shifting visual attention.

In an event-related fMRI (functional magnetic resonance imaging) study, Piazza et al. (2003) found a refined pattern of increasing activation with increasing numerosity by analysing each numerosity separately. Based on a jump in activation between numerosities three and four, the authors argue for a two-process model. Posterier parietal and frontal areas associated with attentional processing started to show activation from numerosity four onwards which led the authors to conclude that attentional processes must contribute to counting, whereas they do not contribute to subitizing. Piazza et al. (2003) propose that this finding confirms the classical parallel pre-attentive - serial attentive dichotomy for subitizing and counting.

---

[1]When using the term "parallel processing", it is often not specified if processing is supposed to occur temporally or spatially in parallel. Studies that adopt the term from visual search studies most likely imply spatially parallel processing. Nieder, Freedman and Miller (2002), however, seem to refer to temporally parallel processing here.

1.3 Theories Explaining a Dichotomy

Several theories have been put forward to explain the discontinuity in enumeration performance and the subitizing phenomenon. Note that all these accounts have been made purely on the basis of the early behavioural studies before evidence from patient, animal and imaging studies had been gathered. However, many of the more recent studies have been explained within the framework of the existing theories and no fundamentally new theory has been proposed on the basis of neuropsychological or imaging evidence.

1.3.1 Pattern Recognition Account

Mandler and Shebo (1982) proposed an explanation of subitizing based on the fact that small numbers of items are more likely to form a familiar pattern than larger numbers. For example, 2 items always form a line, 3 items almost always a triangle. With 4 or 5 items, there are more possible combinations and fewer highly familiar patterns can emerge. The situation worsens with even more than 5 items. Subitizing therefore was supposed to rely on the fast recognition of familiar patterns: instead of counting 3 items, the recognition of a triangle indicates that there must be 3 items present. Mandler and Shebo (1982) argued that the recognition of such canonical shapes is a learned mechanism which may explain why adults have a slightly larger subitizing range than children. They did not make a specific claim about the allocation of attention in pattern recognition.

Several criticisms can be made on this account. First, the data on which Mandler and Shebo (1982) base their hypothesis is far from unequivocal. They demonstrated that subjects are much better at determining the number of dots in regularly arranged dot patterns (such as an equilateral triangle or square) than in randomly arranged dot patterns. However, this was only true in the counting range, whereas in the subitizing range, subjects were equally fast and accurate for both types of arrays. Mandler and Shebo infer from this result that the subitizing process must be the same as the pattern recognition process as there is no difference in performance. However, even with random arrays, subjects performed at ceiling in the subitizing range (with almost no errors and fast reaction times) and thus an effect of regular arrangement

could not have been observed. A particular spatial arrangement of dots might well have an effect in the subitizing range if the task is rendered sufficiently difficult, though this was not established. Thus, the evidence supporting a pattern recognition hypothesis is weak.

Second, subitizing occurs as well when items are all arranged in a linear array and do not fall into regular patterns (Atkinson, Francis & Campbell, 1976). Third, the hypothesis only holds true for arrays of 3 items, possible patterns with four items are much more numerous, unpredictable and can even be ambiguous (e.g. 4 items can form a triangle). Fourth, the theory does not explain well subitizing of targets among distractors. For example, recent evidence by Alston and Humphreys (2004) showed that subitizing of static targets is impaired when they are surrounded by rigidly moving distractors. Pattern recognition theory cannot account for this result: it is not clear why recognition of a static pattern should be impaired due to the mere presence of moving distractors.

1.3.2   Spatial Frequency Account

Based on their finding that subitizing is impaired when items are positioned too close together, Atkinson, Campbell and Francis (1976) suggest that the neuronal populations responsive to numerosity might be the same as or related to those responsive to low spatial frequency. This idea is based on neurophysiological findings in animals showing that low spatial frequency channels have excitatory and inhibitory side bands that respond to a certain number of contrast changes in a stimulus, corresponding to up to 4 items (e.g. 4 bars in a striped array). If numerosity processing is mediated by low spatial frequency channels, then this would explain both the observed limit of 4 items as well as the observed impairment to subitize at high spatial frequencies.

However, this account faces a similar problem than the pattern recognition account: it does not explain enumeration of targets among distractors. When distractors are present, the information from low spatial frequency channels must be filtered through a stage with target-distractor segregation, otherwise it provides misleading information. Atkinson, Francis and Campbell (1976) also propose that perceptual grouping might be a crucial process related to enumeration, which is an interesting idea and will be discussed more in detail in Chapter 2.

1.3.3   Short-term Memory Account

Klahr (1973) was the first to suggest that the discontinuity in enumeration slopes may be explained by the limited capacity of visual short-term memory (VSTM). Many studies have demonstrated a capacity limit in visual short-term memory at about 4 items and performance slopes have very similar characteristics than those in enumeration (e.g. Luck & Vogel, 1997). Cowan (2001) echoed the argument from Klahr (1973) and related short-term memory to attention. He assumed that it is the focus of attention that has a limited capacity and that determines how much information is stored in short-term memory. According to Cowan (2001), visual short term memory has a fixed number of slots, the "magic number 4". Following his logic, subitizing is attentive as only information within the attentional focus is encoded in VSTM. If items to be enumerated fall within the attentional focus, then up to 4 can fill the slots of the memory store without attentional costs, i.e. more or less in parallel. As the number of memory slots are fixed to four, the store needs to be emptied and refilled at higher numerosities. Conceptually, this account is very similar to the FINST account which is described below.

1.3.4   FINST Account

One of the most influential and extensive theories of enumeration has been put forward by Lana Trick and Zenon Pylyshyn (1993; 1994) who were the first to relate the enumeration process with theories of visual attention.

They were also the first to manipulate the stimuli's visual features in their experiments such that the discontinuity between subitizing and counting disappeared. For example, reaction times when enumerating concentric squares followed a continuous linear function throughout all numerosities, whereas enumerating spatially separated squares produced the familiar kink in the curve. Furthermore, they designed displays with classical Treisman conjunction and disjunction stimuli. When subjects were asked to enumerate Os amongst Xs (the classical "pop-out" search), a discontinuity in reaction time performance was found, but when Os are imbedded amongst Qs (the classical "attentive" search), no such discontinuity was observed. Similarly, when Trick and Pylyshyn presented subjects with white lines amongst black lines (or horizontal lines amongst

vertical lines) reaction time data demonstrate a subitizing - counting discontinuity (the disjunction condition), whereas no such discontinuity was observed when subjects enumerated white vertical lines amongst white horizontal and black vertical lines (the conjunction condition). This suggests that when serial attentive processing is required to distinguish targets from distractors (following Treisman's claim), there is no evidence for two separate enumeration modes.

Trick and Pylyshyn base their theory of enumeration on these findings and specifically argue how the discontinuity relates to attention. The theory is grounded on Pylyshyn's FINST hypothesis (1989), which puts forward a visual indexing mechanism for item individuation. After a pre-attentive process of feature detection and grouping, "Fingers of INSTantiation", a sort of reference token for the spatial location of items, is assigned to each item ("items" are defined here as feature clusters). This process allows the visual system to individuate each item and treat it separately, which Trick and Pylyshyn propose is crucial for deriving the numerosity of the set. According to the theory, this indexing stage occurs pre-attentively and spatially in parallel and is followed by a serial attentive stage where the parts of an object are integrated into a whole and finally matched to a memory representation. Crucially, the FINST stage is hypothesized to be of limited capacity, that is, only a limited number of index tokens are available, in this case about four. As in the memory-account, this limit explains the existence of a subitizing-counting dichotomy. Subitizing can occur quickly and effortlessly, because the numerosity of the display is derived from the index tokens which are assigned spatially in parallel. If there are more items to enumerate than there are tokens, the focus of attention needs to be shifted and tokens have to be reassigned, which results in the more laborious process of counting. However, if the visual properties of the stimuli force the visual system to resolve individual items using spatial attention (as in the experimental manipulations of Trick and Pylyshyn) the fast indexing process cannot be used and items must be serially enumerated.

A few aspects of this account remain questionable. For example, it is unclear when a "feature cluster" is categorised as "item" to be indexed or not. Trick and Pylyshyn state that subjects can choose to assign FINSTs to certain features of the stimuli depending on the task requirements, and are even able to adjust the resolution of individuation. This appears to be in conflict with the idea that FINST indexing occurs pre-attentively. If the amount of available tokens is as limited as hypothesized, the visual system must be very careful in assigning them

to certain features and not to others. Thus, even in a display when targets "pop out" from distractors, some top-down control should be necessary to assign the FINST tokens selectively to the targets and not to the distractors. It is implausible that this could be achieved without some sort of attentional feature categorisation. In fact, Trick and Pylyshyn (1994) make a similar criticism on the memory account arguing that distractors should increase the memory load as they fill up the slots, making the assumption that no selection takes place. If selection between targets and distractors can happen before pre-attentive FINST assignment, it should certainly happen before uptake of items in short-term memory.

In sum, Trick and Pylyshyn believe in a fixed capacity limit and in a functional dichotomy between subitizing and counting but argue that an attentional dichotomy is only present when the items to enumerate can be indexed pre-attentively. If indexing needs attention due to the stimulus properties, then the attentional dichotomy disappears and both subitizing and counting are attentive. However, it should be pointed out that they only suppose the FINST indexing stage to occur pre-attentively, not necessarily the derivation of numerosity. Despite this, their study has been cited several times as evidence of subitizing being a pre-attentive process (e.g. in Wender & Rothkegel, 2000 and in Piazza et al., 2003). Indeed, Trick and Pylyshyn explicitly avoid this claim: "We would like to stress that we are not saying that subitizing is pre-attentive, but rather that subitizing makes use of pre-attentive information. Subitizing is certainly not pre-attentive in the sense that it is automatic and involuntary [...] Furthermore, we are not arguing that subitizing is parallel. [...] we want to argue that subitizing does not occur when serial attentional analysis is needed to resolve and individuate items." (Trick & Pylyshyn, 1994, p. 97 & 98).

1.4   Arguments against a Subitizing - Counting Dichotomy

The idea of functionally and attentionally dissociable mechanisms for enumeration has been questioned on many grounds. For many of the above cited studies, criticism about the validity of the interpretation has been raised, or counter-evidence provided. Furthermore, evidence against pre-attentive processing *per se* has been provided in the visual search literature. In contrast to a dichotomy, the idea of a single continuous enumeration mechanism has been put

forward.

1.4.1 Behavioural Studies

A Problematic Kink in the Curve

Many criticisms have been raised about the conclusions from the behavioural evidence. First, it is questionable whether the observed discontinuity in performance slopes is truly evidence for two functionally distinct processes or whether it is in fact an artefact of curve fitting. Indeed, there is evidence that a bilinear model of the performance data might be too simple. Balakrishnan and Ashby (1991 & 1992) thoroughly assessed both bilinear and exponential models (amongst others) with reaction time data from enumeration experiments and found no evidence for a discontinuity in these reaction time distributions. Instead, they suggest a continuous (though not necessarily linear) increase in mental effort with increasing number of items to enumerate. Another alternative to the bilinear model has been proposed by Nieder and Miller (2004). In the performance pattern of rhesus monkeys, they did not find any sudden change in performance slopes either and modelled the data with a sigmoid function yielding a high goodness of fit score. In retrospect, the data of older studies also appear to have a sigmoid tendency, though this alternative has never been formally tested post hoc.

Second, whether the reaction time slope for subitizing is truly flat and therefore reflecting parallel processing as claimed by Sagi and Julesz (1985) has been questioned. Folk, Egeth and Kwak (1988) followed up on Sagi and Julesz's study and found that reaction time and error rates do in fact increase as a function of target number in the subitizing range. They concluded that there must be a serial component to subitizing. Also Trick and Pylyshyn (1994) reported varying subitizing slopes between 40 and 100 ms per item. According to criteria in the visual search literature, this range of slope cannot be regarded as evidence for pre-attentive processing; reaction time slopes for parallel search have been described as varying between 5 and 30 ms per item depending on the search task (e.g. Wang et al., 1994). This suggests that enumeration involves processes in addition to those in visual search, and that those processes become more time consuming with the number of items to be dealt with. Thus, inferences from visual search

experiments might not also necessarily hold true for visual enumeration. Di Lollo et al. (2001) followed up on Sagi and Julesz's experiments and entirely dismissed the idea that reaction time curves reflect the attentive nature of a task (this study is discussed in more detail in section 1.4.6). Thus, the inference that the relative shallowness of the subitizing slope compared to the counting slope indicates pre-attentive processing is more than questionable.

Third, data from several studies showed that the discontinuity in performance slopes disappears depending on the type of stimuli that are used. As reviewed above, Trick and Pylyshyn (1993) were the first to show that if target-distractor discrimination is attention-demanding (according to the classical interpretation from Treisman's feature integration theory) then the familiar kink in the curve disappears. Although Trick and Pylyshyn explain their data with a fixed capacity-limited indexing mechanism (which cannot be used when attention is needed for target-distractor discrimination), their data suggest that if there was a fixed-limited enumeration mechanism, it would only work for some stimuli but not for others. This is inconsistent with the idea of an ubiquitous, hard-wired "core system" (as suggested by Feigenson, Dehaene & Spelke (2004) or Butterworth (1999)) as it seems to fail for stimuli that are not easily individuated or segregated from other items.

Alston and Humphreys (2004) provided convincing evidence on how the mere presence of certain distractors can influence the efficiency with which items are enumerated. As mentioned above, they found that enumerating moving targets amongst static distractors produced a discontinuity in the reaction time slope, whereas enumerating static targets amongst moving distractors resulted in much higher reaction times, linearly increasing with target number. Naturally, static targets without distractors did exhibit the classical kink. That is, the mere presence of moving distractors makes the enumeration of static items a more difficult and possibly attention-demanding task. Thus, even if there was a specialised subitizing mechanism, it appears to fail easily due to the presence of task-irrelevant distractors.

Weber's Law and Theories based upon it

Surprisingly, many studies have neglected the possibility that the difference between subitizing and counting might be due to the fact that processing load for the brain increases with the

number of items present - and thus that the subitizing phenomenon is nothing more than a ceiling effect. That is, subitizing is a very easy and effortless task due to the little information that has to be processed, whereas counting is more difficult due to the increased amount of items to deal with. The relationship between different intensities, or magnitudes, of stimuli in the real world (e.g. weights, sounds, numerosity) and their representation in the brain has first been described by Ernst Weber. Weber's law states that the ability to discriminate two stimuli from each other decreases with stimulus size (or intensity). In the case of number this means that it becomes increasingly difficult to discriminate two numerosities the bigger they are and the smaller their relative distance or ratio is (also known as the size and the distance effect, respectively). The Weber fraction is defined as the minimum ratio required so that two stimuli can be discriminated successfully (according to a certain criterion, e.g. in 50% of the cases).

Based on this idea, Averbach (1963) was one of the first to note that if the ratio between two numbers is big enough then judgement accuracy and speed would have the characteristics of subitizing. Van Oeffelen and Vos (1982) confirmed this idea experimentally and showed that pairwise discrimination between small numbers (in the case of this study, up to 6) can be done easily as their ratio lies well above the Weber fraction of about .16. They argue that any two numbers whose ratio is above the Weber fraction can be discriminated successfully and conclude that "the idea that the mind can grasp only a small number of objects at once remains quite unsupported by the evidence, if indeed it has any meaning at all." (Van Oeffelen & Vos, 1982, p. 169).

Ross (2003) echoes this argument. By investigating a larger range of numerosities than van Oeffelen and Vos, he found a Weber fraction of .25 that is constant across numerosities. His data show that numbers that are further than 25% apart from each other can be discriminated almost without errors. For numerosities 1 to 4, this is always true as the difference between 3 and 4 is 25%. Numerosities 4 and 5, however, are only 20% apart and thus lie closely below the Weber fraction for succesful discrimination. With higher numerosities, the situation worsens. Thus, based on Weber's law, the study by Ross (2003) provides a straightforward explanation for why small numerosities can be dealt with so effortlessly and why it becomes so much more difficult to handle numerosities of 5 onwards - without the need to postulate two separate enumeration modes.

Gallistel and Gelman argue along the same lines (Gallistel & Gelman, 2000; Cordes, Gelman & Gallistel, 2001). Based on experiments with rats and humans, they propose that the mental representation of magnitude is characterised by scalar variability. This means that the variability with which magnitudes from the real world are represented in the brain is linearly proportional to the size of the magnitude.[2] I.e. the greater the magnitude, the noisier its representation, consistent with Weber's law. In a study where human subjects were asked to press a button a specific number of times under suppression of verbal counting, Cordes, Gelman and Gallistel (2001) found that variability of responses was constant within and outside of the subitizing range. The authors therefore conclude that small numbers are represented on the same continuum as larger numbers, as they are characterised by the same variability signatures as expected from Weber's law. Furthermore, their data showed that this variability signature is typical for a non-verbal enumeration mechanism shared among humans and animals. Verbal counting however, shows binominal variability (decreasing variability with increasing numerosity) and might thus be a different mechanism.

It has be noted that this distinction between non-verbal enumeration and verbal counting is quite crucial and often neglected in enumeration studies. For example, many studies used unlimited presentation times for their numerosity stimuli and recorded reaction times as dependent variable. Even if subjects are prompted to respond as quickly and accurately as possible, it cannot be determined whether they used fast verbal counting or true subitizing (non-verbal perceptual apprehension of numerosity). Verbal counting of small numerosities (or double checking of an already subitized set) will invariably take less processing time than counting larger sets and thus yield similar results than when true subitizing is employed. Given Gallistel and Gelman's hypothesis, verbal counting might be a fundamentally different process than non-verbal apprehension and as such results in some enumeration studies might bear confounds.

Two other related caveats of enumeration studies need to be mentioned. Related to Weber's law is the fact that more stimuli bear more processing demand, i.e. numerosity correlates with task difficulty. Thus it is difficult to distinguish whether higher reaction times and lower accuracy in the counting range reflect a qualitative difference in enumeration or simply higher processing demands. In order to circumvent this problem, variable numbers of distractor are

---

[2]Others (e.g. Dehaene & Changeux, 1993) have argued that this relationship is logarithmic instead of linear.

useful to de-correlate the number of items to enumerate with the overall processing demand. Many of the older studies did not use distractors and presented their stimuli on an empty background (e.g. Mandler & Shebo, 1982; Atkinson et al., 1976a & b), and thus might bear a confound. On the other hand, however, using distractors may impose an additional processing demand as targets need to be segregated from distractors, and depending on how difficult this segregation is, enumeration performance can be severely affected (cf. Trick & Pylyshyn, 1993; Alston & Humphreys, 2001). This issue is addressed in detail in Chapter 2.

If variable amounts of distractors are not used, another confound can occur. With increasing numerosity, several other parameters of the stimulus also increase, such as the area covered, density and overall luminance. These covarying cues can be used to to make a quantitative judgement about the display, in addition or instead of numerosity. In order to prevent this confound, displays must be adjusted to control for these parameters, for example, by keeping the covered area constant while varying the numerosity. Replication of some older studies revealed that when such continuous parameters are controlled for, the effect of numerosity disappears (e.g. in infant studies; Xu, 2003). It is, however, unavoidable that at least one visual parameter will positively correlate with numerosity. If no distractors are used, the best solution is provided by Nieder and Miller (2004) who mixed several types of displays in which at least one visual parameter stayed constant despite increasing numerosity.

1.4.2   Limited Capacity in Other Fields

The question of the existence of a fixed capacity limit is also discussed in the field of visual-short term memory and attention. Whereas some argue for a fixed capacity limit of 4 items in visual short-term memory (Luck and Vogel, 1997; Zhang & Luck, 2008), others provided evidence against such a limit. Alvarez and Cavanagh (2004) showed that capacity varies between 1.6 and 4.4 items depending on stimulus complexity. Also Bays and Husain (2008) recently demonstrated that the number of items held in VSTM depends on the precision with which these items are encoded.

Also in the field of attention it has been argued that the number of locations that can be selected at once varies with the precision which is required by the task (Franconeri,

Alvarez & Enns, 2007). A similar case has been made for the number of objects that can be attended, it does not seem to be restricted by a fixed limit, contrary to Cowan (2001). Davis et al. (2001) kept the size and overall complexity constant while varying the number of items and demonstrated that it is not a fixed number of objects that poses the limits of attention, but the complexity of the stimulus and thus the perceptual processing demand. A similar argument has been made by Alvarez and Franconeri (2007) for the case of multiple object tracking. By varying the speed and spatial distance of moving items to be tracked, they demonstrated that as little as 1 and as many as 8 items can be tracked and that the limit of tracking is related to the spatial resolution of attention. This speaks for attention being a flexibly allocated resource rather than being restricted by a fixed capacity limit.

Taken together, all these findings are in contradiction with the proposal by Cowan (2001) of a fixed number of objects that can be attended to and stored in visual short-term memory and thus give rise to the subitizing phenomenon.

1.4.3   Neuropsychological Studies

Vuilleumier and Rafal (2000) claim that neglect patients can enumerate items in their neglected field although being unable to discriminate and locate them. The authors take this as evidence for a pre-attentive subitizing mechanism. However, it has to be pointed out that patients were very error-prone in distinguishing 1 from 2 items in the neglected (contralesional) field even when nothing was present in the intact (ipsilesional) field. The authors admit that this finding conflicts with a wholly pre-attentive mechanism because such a mechanism should be preserved within and across hemifields.

However, the implications of this conflicting result might be more profound than addressed by Vuilleumier and Rafal (2000). Given the fact that the enumeration task included only 3 choices of response (1, 2 or 4 items distributed across both hemifields), this result points to the possibility that patient's judgement was categorical instead of numerical. For example, in the case of numerosity 4, two items were presented in the intact and two in the neglected field. If patients correctly detected two items in the intact field and "something" in the neglected field, they would be correct in responding 4 even if they could not distinguish the items in the

neglected field. As response choices did not include numerosity 3, it could not be determined whether patients were able to distinguish 2 from 3 or 3 from 4 if some of the items were displayed in their neglected field. Therefore, it is questionable whether neglect patients are truly able to subitize in their neglected field, thus this argument for a pre-attentive subitizing mechanism is weakened.

1.4.4   Infant and Animal Studies

With regard to small numerosity discrimination in infants, the effect disappears under controlled experimental conditions. Xu (2003) controlled the visual display for continuous variables (such as surface area, item size and circumference) and found that 6-month olds are unable to discriminate arrays of 2 versus 4 items although they succeed with larger collections differing by the same ratio (4 versus 8). These results show that when discrimination needs to be done solely on the basis of numerosity rather than covarying variables, the case for an innate ability to discriminate numerosities in the subitizing range, as originally claimed by Starkey and Cooper (1980), is weakened. [3] However, given that judging smaller collections should pose an easier task, these results are rather counterintuitive, particularly with respect to findings in non-human primates which generally found that discrimination becomes harder at larger collections (e.g. Hauser, Carey & Hauser, 2000).

However, also the animal literature does not unequivocally argue for an innate, hard-wired system dedicated to small numerosities. In the case of rhesus monkeys, the set size limit has been found to be not entirely fixed to 3 or 4 items, as these animals can also spontaneously represent higher numerosities (i.e. 5 - 9. Brannon & Terrace, 2000).

Based on the behavioural and neural findings of their monkey studies, Nieder and Miller (2004) follow in principle the argument by Gallistel and Gelman (2001) by proposing an analog magnitude system, a number estimation system that becomes systematically less precise with increasing numerosity. Both their behavioural data as well as the tuning functions of numerosity responsive neurons exhibited a clear Weber fraction signature. In contrast to

---

[3]It should be noted that both in infant as well as in animal studies, it is usually the discrimination of two numerosities from each other that is studied rather than classical subitizing.

Gallistel and Gelman, however, Nieder and Miller argue for a logarithmic relationship between real magnitudes and their mental representation. They also argue for a rather smooth transition between low and high numerosities (following a sigmoid function) rather than a clear cut-off point. Additionally, also Nieder and Miller note that increasing reaction time while judging numerosity might be related to task difficulty and detection processes rather than numerical coding per se.

1.4.5   Imaging Studies

The evidence from imaging studies is not as clear-cut as the interpretations often suppose. In a PET study, Piazza et al. (2002) did not find any difference in brain activation between subitizing and counting. Their results showed a pattern of activation that is similar in all enumeration ranges with an overall increase of activation with the number of items in the display, a finding that argues against separable neural systems. [4] Furthermore, two arguments can be made about the earlier reported imaging studies that argued for a subitizing-counting dichotomy (Sathian et al., 1999; Piazza et al., 2003). First, when the two enumeration modes are contrasted, the imaging data is, by hypothesis, averaged across subitizing trials and across counting trials. It might not be surprising to find more areas activated in the counting range than in the subitizing range when applying this method. If a parameter varies along a continuum, dividing both ends of the continuum into two different data pools will very likely result in a difference in the means of the pools. Thus, the observed difference in activation between subitizing and counting trials might be an artefact of averaging. Second, as mentioned earlier, processing demand increases with increasing numerosity and therefore might likely require a more widespread brain network, a possibility that is not taken sufficiently into account by these studies.

A recent fMRI study did not directly compare subitizing and counting range, but contrasted number comparison for numerosity ranges 1 to 4 and 10 to 40 (Ansari et al., 2007). Small compared to large numerosities activated the right temporo-parietal junction, an area associated with stimulus-driven attention (Corbetta & Shulman, 2002). This is the first evidence

---

[4]In their subsequent fMRI study, Piazza et al. (2003) do not give a plausible account of how their interpretation of the fMRI results is reconciled with the findings from this earlier PET study. In their fMRI study from 2003, they argue for separate neural correlates for parallel pre-attentive subitizing and serial attentive counting, (see section 1.2.4).

that a higher-level area involved in attentional processes might play a role in small numerosity judgement, thus speaking against pre-attentive subitizing. As activations for large compared to small numerosities did not yield the same activations, Ansari et al. (2007) concluded that small numerosities might be processed differently from larger numerosities, thus following the idea of a functional dichotomy.

1.4.6   Arguments from the Visual Search Literature

The classical dichotomous model of a parallel and pre-attentive processing stage (implemented by specialised hard-wired analysers) and a serial, attentive stage as originally suggested by Treisman and Gelade (1980) or Julesz (1984) has long been questioned in the visual search literature. Many scientists nowadays agree that a bimodal model is an overly simplified account of attentional processing (e.g. Duncan & Humphreys, 1989; Nakayama & Joseph, 1998). For example, after reviewing a vast amount of literature, Driver et al. (2001) concluded that "... the attentional state of the observer can modulate *many* levels of perceptual processing, rather than kicking in only at a specific 'attentive' stage, following an initial 'preattentive' state." (Driver et al., 2001, p. 64).

Norman and Bobrow (1975) were one of the first to put forward a very interesting approach from information theory explaining how performance and limited processing resources interact - without invoking a two-stage model. They supposed that performance can be limited by two factors: a sensory data limit (the amount of information reaching the brain) and a resource limit (the amount of attention, memory or processing effort that can be allocated to the task). If sensory data is not limited, i.e. stimulus strength is well above detection threshold, then the amount of resources is the only limiting factor of performance. The relationship between resources and performance is depicted in Fig. 1.2a. In an easy task (e.g. a typical "pop-out" visual search task), maximum performance is reached already with a very small amount of resources. In a difficult task (e.g. a typical conjunction search task), more resources are needed to gain maximum performance. Importantly, Norman and Bobrow (1975) made predictions about how performance is affected by dual task situations: if two tasks compete for the same resources, and one belongs to the easy category (i.e. reaching performance asymptote with very

few resources) then plenty of resources can be allocated to the second task, where performance maximum is quickly reached. If, however, one of the two tasks requires a lot or even the full amount of resources, then performance of the other task cannot reach maximum and is impaired, even if it is an "easy" task. Which of the two tasks get the resources allocated is determined by task instructions. The important implication of this account is that, in the case of visual search, tasks are not divided into two qualitatively separate categories, pre-attentive tasks carried out with in-built analysers and attentive tasks involving higher level processing, but every task being located on different stages on a continuum of attentional resources. Thus, every task requires a minimum amount of resources and the hypothesis of a pre-attentive/attentive dichotomy becomes unnecessary.

Joseph, Chun and Nakayama (1997) provided one of the most striking demonstrations that the concept of pre-attentive processing is flawed - and confirmed the predictions of Norman and Bobrow (1975). By employing a dual-task paradigm they demonstrated that even simple feature detection (orientation odd-ball detection in this case) is severely impaired when an additional, attentionally demanding task is performed at the same time. Consistent with Norman and Bobrow (1975) this shows that even the most "pre-attentive" task can be impaired if the majority of attentional resources is allocated to another task. In a reply to this study, Braun (1998) showed that expert subjects are much less affected by dual-task conditions than novices. However, this result can also be nicely explained within the framework of Norman and Bobrow (see Fig. 1.2 b). Training shifts the performance by resources curve further to the left, such that fewer resources are needed for successful performance. This implies that a resource-demanding additional task may have little or no impact when one or both of the tasks are well trained (Joseph, Chun & Nakayama, 1998).

Also Di Lollo et al. (2001) dismissed the idea of a privileged, pre-attentive processing stage. They followed up on Sagi and Julesz's study (1985) and, by imposing dual-task conditions, were able to selectively alter the performance slopes of both subitizing and discrimination tasks. Based on these results, they argued that reaction time slopes do not indicate the nature of a task but whether it involves a task switch. They propose a high-level input filtering system that can be dynamically reconfigured, depending on task demands, to manage the processing of stimuli efficiently. When the task involves a task switch, the system needs to be reconfigured which

a

P_{max}

Performance

Easy Search

Hard Search

R_{easy} R_{hard} R_{max}

Resources

b

P_{max}

Performance

After Training

Before Training

R_{post} R_{pre} R_{max}
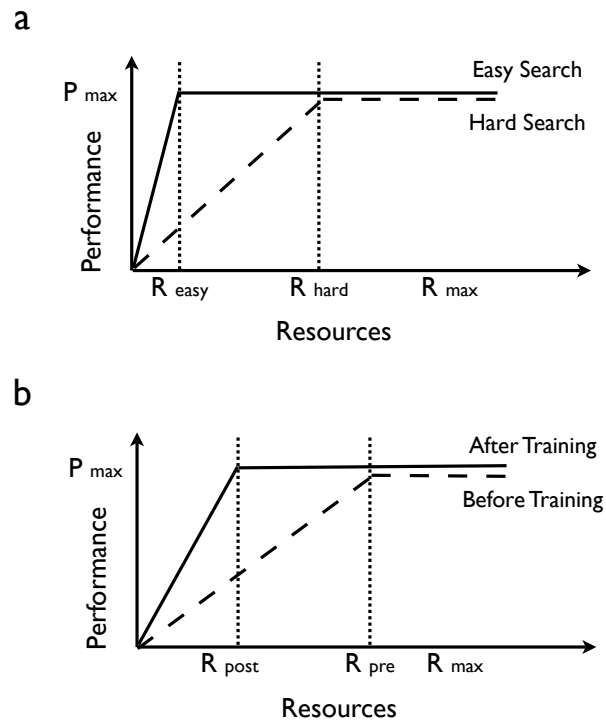
Resources

Figure 1.2: Norman and Bobrow's Model

Relationship between performance and processing resources as hypothesised by Norman and Bobrow (1975). a) In an easy search task, maximum performance (P_{max}) is achieved with relatively little resources, whereas hard search requires far more resources to reach the same level of performance. b) The effect of training. A hard task requiring a lot of resources before training will take up much fewer resources after training.

takes up processing time and results in steep reaction time curves. This idea is in line with Wolfe's "guided search" theory (Wolfe, 1994) in which the efficiency of search is determined by higher level brain areas modulating the processing of lower levels. Both concepts circumvent the necessity of a hard-wired pre-attentive module and argue for a unimodal, continuous mechanism for the allocation of attention.

1.5   Objective of this Thesis

As reviewed, the evidence for the traditional claim of two functionally and attentionally distinct enumeration mechanisms for small and large numerosities is very controversial. It appears that the recent findings from the visual search literature have not been sufficiently taken into account in the numerical cognition literature. This calls for a systematic reconsideration of the theories concerning visual enumeration.

The question addressed in this thesis was whether subitizing is a truly parallel and pre-attentive process or whether it might require visual attention. Given the evidence from the visual search literature that even feature-based detection of a single odd-ball can be compromised by reducing the amount of attentional resources allocated to the task, it appears unlikely that the detection of several odd-balls should occur pre-attentively. On the other hand, a considerable amount of enumeration literature shows that the subitizing phenomenon is a very robust finding and difficult to disrupt experimentally.

The focus of this thesis lies on the attentional aspects of subitizing, and is thus less concerned with the existence of a functional dichotomy between subitizing and counting. However, as the hypothesis of pre-attentive subitizing is one of the key arguments for the special status of the subitizing mechanism, arguments for or against pre-attentive subitizing will necessarily affect the interpretations towards the existence of an attentional as well as a functional dichotomy. Furthermore, emphasis is given on the perceptual apprehension of small numerosities rather than on the mechanisms of verbal counting. Most experiments included numerosities from the counting range, however, mainly to investigate whether the effects of experimental manipulation also hold for the estimation of higher numerosities. Note that also with those higher

numerosities, enumeration was studied in the sense of perceptual apprehension, i.e. estimation rather than verbal counting.

In this thesis, pre-attentive subitizing was not sought to be determined by analysis of performance slopes, nor were slopes formally examined for any discontinuities. In the above sections, it is discussed in detail how many confounds an inference from performance slopes bear, both with respect to attentional requirements of the task and with respect to whether they reveal a unimodal or a bimodal process. Instead, attentional set and attentional resources allocated to the enumeration task were manipulated experimentally and the effects of such manipulations were investigated in detail. At the time of the start of this project, this had never been attempted before. Trick and Pylyshyn's manipulations of using stimuli with differential attentional demands were the only direct attentional manipulation that was implemented in an enumeration task.

In line with the classical visual search literature (Treisman & Gelade, 1980; Julesz, 1984), a pre-attentive process is defined here as a process independent of the available amount of attentional resources as it is supposed to occur *prior* to any allocation of attention. Furthermore, within such a pre-attentive processing stage, stimulus attributes (such as colour, motion or perhaps numerosity) are extracted automatically and spatially in parallel. In the case of subitizing, this means that small numerosity information is retrieved from a visual scene for possibly several sets of items in parallel, automatically (i.e. involuntarily and without any top-down control) and without the necessity of allocation of attention. This definition predicts that subitizing performance should be unaffected by attentional set and by the amount of attention available to the enumeration task. This hypothesis was tested in the present work.

1.6   Overview of the Experiments

Three main experiments are presented in this thesis. The first focussed on the claim of a spatially parallel subitizing mechanism and investigated enumeration performance when feature-based attention could be drawn to one subset of items compared to when it had to be distributed to two subsets. A cueing paradigm was established where subjects were cued to the relevant subset

either before or after stimulus presentation. By manipulating the perceptual characteristics of the stimuli, the impact of perceptual grouping on enumeration was studied and its relationship to item discrimination and visual search.

The claim of a pre-attentive subitizing mechanism was furthermore tested by reducing the attentional resources available to the enumeration task. Experiment 2 established a dual task paradigm where the allocation of such resources was systematically modulated. In addition to imposing a secondary task, the processing demands and thus the attentional load of this additional task were manipulated. The effect of reduced attentional resources was also investigated for single item detection in comparison to enumeration.

Lastly, in Experiment 3 the interaction between enumeration and attentional load was investigated on the neural level using fMRI. The dual task paradigm from Experiment 2 was adapted for fMRI and neural activations for enumeration with full attentional resources were compared to activations related to enumeration with reduced attentional resources.

Chapter 2

Experiment 1: Exploring the Effects of Attentional Set on Subitizing: a Cueing Paradigm

'

2.1 Introduction

The motivation for the paradigm presented here arose from the fact that, in previous studies, it was always unambiguous which set of items should be enumerated. Either no distractors were used (e.g. Mandler & Shebo, 1982) or the targets to enumerate were clearly defined (e.g. Trick & Pylyshyn, 1993). In the real world however, this is a rather rare situation. Looking at a table, for example, opens up many different ways of enumerating items. One could count the number of glasses, or the number of filled and empty glasses. The number of forks and knifes could be enumerated separately, or every item of cutlery. The idea of automatic and pre-attentive subitizing implies that the numerosity from every small set in a visual scene should be extracted in parallel. It appears more likely, however, that in most cases the task demands (e.g. are there enough forks for everyone?) require that attention is focussed on just one subset and it would

be a waste of resources to enumerate every possible subset in parallel. The question now arises: is numerosity extracted from every possible subset automatically and attention just selects the required response, or is subitizing under top-down control and only the subset of interest is enumerated?

The current cueing paradigm was established to create a situation where it was unambiguous which items would be enumerated and attention could be allocated to the relevant subset (pre-cue condition). This was compared to the situation when it was ambiguous which of two subsets would be enumerated and attention had to be distributed to both sets (post-cue condition). Thus, this comparison comprised a manipulation of attentional set and was deemed suitable to test the hypothesis of a parallel and pre-attentive subitizing mechanism. Furthermore, it was investigated whether manipulation of attentional set interferes with the enumeration process *per se* or with any of the processes enumeration might be contingent on, i.e. item discrimination and perceptual grouping. Experiment 1.1 established the cueing paradigm, whereas in Experiment 1.2 the features by which items are grouped were altered in order to investigate the role of perceptual grouping. Experiment 1.3 tested how differential grouping features affected item discrimination. Lastly, Experiment 1.4 investigated the effect of attentional set on a parallel visual search task in comparison to enumeration.

2.1.1   The Paradigm

Attentional set was manipulated by presenting two subsets of stimuli simultaneously (e.g. black and white dots intermixed in the display) and by cueing subjects to the relevant target subset either before or after stimulus presentation. In the pre-cue condition (presenting the cue before the stimulus), subjects were able to allocate their attention to the relevant subset and ignore the distractor subset. In the post-cue condition (presenting the cue after the stimulus), however, subjects could not selectively attend to one subset but had to spread their attention to both subsets. In order to make a correct judgement about the target subset, subjects therefore had an attentional advantage in the pre-cue condition which was eliminated in the post-cue condition. If the required judgement is a truly pre-attentive process, the slight advantage of focussing attention to the relevant subset should not matter. Furthermore, if the judgement involves a

truly parallel process, the information of both subsets should be available in parallel and pre-cueing should not present an advantage, but rather merely indicate what shall be reported. Consequently, it was predicted that behavioural performance should not differ between the two cueing conditions if the judgement relies solely on pre-attentive and parallel processing. If there is a difference, however, then at least some of the processes involved in the judgement cannot be performed in parallel and/or require focussed attention.

Note that as the stimuli of both subsets were spatially intermixed, this manipulation tapped into feature-based rather than spatial attention. Furthermore, attentional set was varied between focussed attention and distributed attention (rather than attended versus unattended, for example).

The current paradigm is in some respect similar to Sperling's classical partial/whole report procedure (Sperling, 1960). Under whole report conditions, subjects are asked to report all information from a briefly presented display whereas under partial report, subjects only retrieve a subset of information to which they are cued shortly after stimulus presentation. Typically, subjects' performance is better in partial than in whole report conditions, particularly if the capacity limit of visual short-term memory is exceeded for whole report but not for partial report (Sperling's original experiment investigated information retrieval from iconic memory). Wender and Rothkegel (2000) applied a whole/partial report procedure to enumeration by asking subjects to report the numerosity of one of two simultaneously presented dot patterns (partial report) or the numerosity of both patterns separately (whole report). Dot patterns were separated into two subsets either spatially or by form (circles versus triangles) or both. In line with Sperling (1960), partial report performance was consistently better than whole report performance. More interesting, however, was the fact that partial report performance dropped dramatically when stimuli had to be separated by form compared to by spatial distribution. This indicates that grouping by form interferes with enumeration whereas grouping by space does not, or at least to a lesser extent (unfortunately, Wender and Rothkegel (2000) provide insufficient statistics on their data). The criterion by which subsets are defined therefore seems to play a crucial role in this sort of paradigm, an issue that is addressed in Experiment 1.2.

Applied to the current experiments, both cueing conditions correspond to a partial

report (both Sperling (1960) and Wender & Rothkegel (2000) employed a post-cue) whereas the control conditions of Exp. 1.1. and 1.2. (enumeration of all items presented) correspond to a whole report. On the basis of the studies by Sperling and Wender and Rothkegel, the order of the cue with respect to the stimulus should not matter regarding the encoding in visual short-term memory (however, the lag between stimulus and post-cue does matter as shown by Sperling, 1960). Pre-cueing compared to post-cueing therefore should rather influence *which* information is encoded in visual short-term store and this should be determined by attention (e.g. Duncan & Humphreys, 1989). Therefore, the reported experimental manipulation should tap into an attentional rather than a mnemonic capacity-limited process.

2.2 Experiment 1.1 Effects of Attentional Set on Subitizing

2.2.1 Rationale

The aim of this experiment was to investigate subitizing performance under differential attentional set in order to test the hypothesis of a parallel and pre-attentive subitizing mechanism. In this task, a display of randomly distributed white and black dots was briefly presented and subjects were given a cue to the relevant feature (black or white) either before (pre-cue condition) or after (post-cue condition) stimulus presentation. Following the above described logic, it was predicted that subitizing performance (measured as accuracy, reaction times and response deviation) should not differ in both conditions if subitizing is a pre-attentive and parallel process. However, if any of the processes involved in retrieving numerosity from the target subset require attention or cannot be performed in parallel, performance should be impaired in the post-cue condition compared to the pre-cue condition.

The task was complemented by a control condition in which subjects enumerated the whole set of dots on the screen without dividing them into targets and distractors. This condition was included for two reasons: First, subjects baseline enumeration performance under the specific task conditions was established, a condition in which no target-distractor segregation was required. Second, the effect of feature homogeneity or heterogeneity on enumeration performance was investigated. That is, in half of the trials dots were homogeneous (either only black

or only white), whereas in the other half, dots were heterogenous (black and white). Feature heterogeneity (in this case luminance heterogeneity) could have two different effects: subjects could use the heterogeneity to their advantage in enumeration by segregating the whole set into two subitizable set and employing a subitize-and-add strategy. For example, 3 white dots and 4 black dots could be subitized in parallel and added up to a more precise or faster result than enumerating 7 white dots. In this scenario, increased accuracy and/or reaction times in the heterogenous trials compared to the homogenous trials would be expected. Alternatively, it could be argued that feature heterogeneity adds irrelevant perceptual information to the task, taking up processing resources and leading to increased reaction times or error rates compared to homogenous trials.

If stimulus exposure is unlimited or sufficiently long (as mostly employed in previous studies), it is impossible to distinguish whether subjects truly subitized or verbally counted in the small numerosity range. As this experiment was primarily designed to test the ability to immediately apprehend the numerosity of a subset without verbal counting, counting was deliberately prevented by employing short stimulus presentation times (between 50 and 150 ms) and a mask. Short presentation times had the further advantage of preventing subjects to make extensive eye movements and to scan through the display. Stimulus presentation time was varied in order to investigate whether the effect of attentional set changes depending on exposure duration.

As counting was prevented, it would be more appropriate to speak of "estimation" or "approximation" for the enumeration of higher numerosities rather than "counting". For the sake of simplicity, however, the numerosity range above 4 will in the following be referred to as the "counting" range in line with the traditional terminology. This shall not imply that any true counting has taken place.

As the process of subitizing was the main focus of the experiment, the cued conditions consisted of many more trials in the subitizing range than in the counting range. Counting trials were included as catch trials: to reduce the probability of guessing and to prevent subjects from categorising subsets automatically as distractors when they exceeded the subitizing range. The control condition, however, consisted of equal number of trials for each numerosity 1-9.

Perceptual biases towards one subset were avoided by gamma-correcting the monitor for output luminance and adjusting the grey background such that both black and white dots were made as equally salient as possible.

2.2.2   Methods

Subjects

34 subjects (mean age: 25.3, range: 21-38, 28 females) were tested in total, 21 with 50 ms, 7 with 84 ms and 6 with 150 ms stimulus presentation times. Subjects had normal or corrected to normal vision, provided informed consent and were paid for their participation. They were tested in a quiet and dimly lit room, seated at 50 cm distance from the computer screen using a chin rest. Due to the short stimulus presentation times, subjects were only tested when they reported to be able to perceive the stimulus and to distinguish black and white dots.

Stimuli

The stimulus consisted of black and white dots (diameter of $1°$ visual angle) drawn on a grey background. The background was gamma-corrected for output luminance and adjusted to mid-grey between black and white. The x and y coordinates of each dot position were randomly assigned anew for each trial to avoid regular patterns. No grid was used, formation of regular patterns could have therefore only occurred by chance. Minimum distance between dots was one dot diameter. Dot spreading was restricted to a field of $16° \times 12°$ at the centre of the screen (half of the screen width and half of its height) to avoid stimulus presentation too far in the periphery.

The mask was generated anew for each trial with 1000 black and 1000 white dots, partly overlapping and randomly distributed over the whole screen. The cue consisted of the word "BLACK" or "WHITE" displayed in either black or white, respectively. The fixation cross was red to prevent invalid perceptual cueing from the cross colour. Stimuli were created using the Cogent and Cogent Graphics toolboxes (www.vislab.ucl.ac.uk/Cogent/) for MATLAB

(Mathworks, Inc.) and presented on a CRT monitor (14' Sony Multiscan 110ES, 60 Hz, 1024 x 768 resolution).

Task and Experimental Procedure

The task comprised three experimental conditions. In the control condition, subjects reported the total number of dots displayed on the screen, regardless of their colour. In the pre-cue condition, subjects selectively enumerated those dots whose colour was announced in the cue preceding the stimulus presentation. The post-cue condition required the same, with the difference that the colour cue was displayed after stimulus and mask presentation.

In the control condition, each trial consisted of a centered fixation cross displayed for 1000 ms, followed by the stimulus for either 50, 84 or 150 ms, respectively, and subsequently the mask (Fig. 2.1 a).[1] The mask stayed on the screen until subjects responded and was then followed by the next trial. In the pre-cue condition, the cue was displayed for 500 ms after the fixation cross, followed by another fixation cross for 500 ms and subsequently followed by the stimulus and the mask (Fig. 2.1b). The mask remained on the screen until subjects responded. In the post-cue condition, subjects received the colour cue after presentation of the stimulus and 500 ms of mask. Here, the colour cue stayed on the screen until response (Fig. 2.1c). Subjects always performed first the control condition, followed by the cued conditions, with the order of the cued conditions counterbalanced across subjects (with exception to the group tested with 84 ms stimulus presentation time (7 subjects) who performed the task in the order control - precue - postcue).

Each condition was preceded by an instruction screen and subjects were prompted to take a break after half of the trials in each condition. About 20 practice trials were given for each condition, depending on how quickly subjects felt confident about the task. Subjects were instructed to be as accurate as possible and to give their best guess when they were unsure. Emphasis was given on accuracy rather than speed. Subjects responded via the number keys 1 - 9 on the computer keyboard after each trial. No feedback was given. Reaction times were

---

[1]Given the refresh rate of 16.6 ms on a 60 Hz monitor, the actual presentation time was 83.3 ms. However, to ensure that the stimulus was presented for full 5 refresh rates, the MatLab programme was set to 84 ms. In the following, this experimental condition will be referred to as the 84 ms condition.
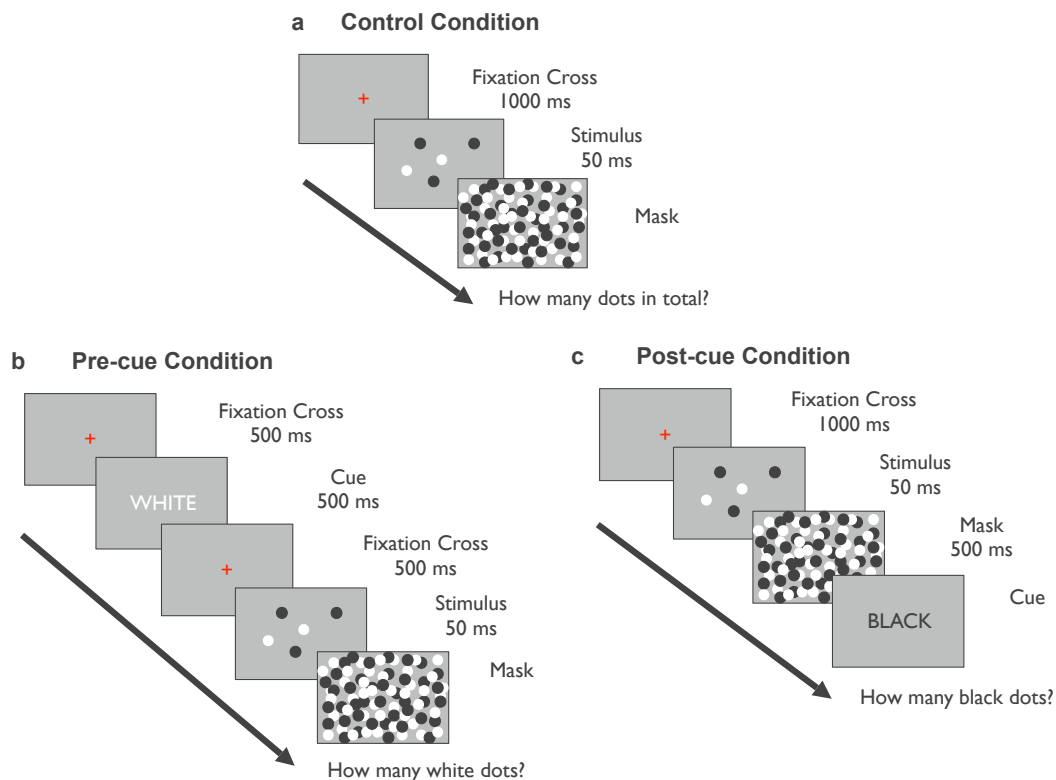
Figure 2.1: Experimental Procedure of Exp. 1.1.

a) Control condition b) Pre-cue condition c) Post-cue condition. This is a schematic representation, stimuli are drawn in a different scale than in the real stimulus.

recorded from onset of the stimulus until response, as well as accuracy and response number. The whole experimental session lasted for 45 - 60 min.

Experimental Design

Each condition consisted of 144 trials, with the total number of dots counterbalanced in each condition. In the control conditions, total dot number ranged from 1 to 9 and was equal to the number of dots to enumerate (target number), therefore each numerosity 1 - 9 was displayed 16 times. Half of the trials in the control condition were heterogeneously coloured (black and white, number of each counterbalanced), half of them were homogeneously coloured (half of those trials were only black, the other half only white). In the cued conditions, the number

of dots to enumerate was not equal to the total number of dots displayed (ranging from 3 -
10, presented 18 times each), as subjects were requested to report the numerosity of just one
subset. The frequency of each target set numerosity therefore varied across the condition, and
was designed so that a total of 112 trials required a correct response of either 1, 2, 3 or 4 (28
trials each) and 32 trials required a correct response of 5 up to 9 (exact trial frequencies: 12 trials
at target numerosity 5, 8 trials at numerosity 6, 6 trials at numerosity 7, 4 trials at numerosity
8 and 2 trials at numerosity 9). All trials in the cued conditions were heterogeneously coloured,
half of them required to report the white subset, half of them the black subset. The order of
trials was pseudo-randomised for each subject and each condition anew.

Data Analysis

Mean accuracy, mean reaction times, mean responses, mean response variances and response
deviation were calculated for each target number and each subject separately and then averaged
across subjects. In the post-cue condition, reaction times were corrected for the 500 ms the mask
was displayed on the screen. Repeated measures ANOVAs were employed throughout. Within-
subject factors were experimental condition (3 levels for control, precue and postcue) and target
number (number of levels depending on which range was investigated). Between-subject factor
was stimulus presentation time (3 levels for 50 ms, 84 ms and 150 ms). Analyses were carried
out for all data pooled together and for each presentation time and numerosity range separately.
Based on the experimental hypothesis, planned comparisons (also repeated measure ANOVAs)
were employed between the pre-cue and post-cue condition. In all other post-hoc comparisons,
α levels were corrected for multiple comparisons after Bonferroni. In order to determine whether
accuracy performance was significantly above chance level, one sample T-tests were employed.
The same tests were employed to determine whether mean responses differed from the correct
response.

Chance level was defined on the basis of the trial frequencies for each target numerosity
(i.e. number of trials for a certain target number divided by the total number of trials in each
condition). The control condition required responses from 1 to 9 equally distributed across
all trials, i.e. chance performance was at a ninth (0.11) for each target number. Given the

differential trial frequencies in the cued condition, chance level was at .194 for target numbers 1 - 4 each, .083 for target number 5, .055 for 6, .042 for 7, .028 for 8 and .014 for target number 9. The possibly problematic issue of defining chance performance in enumeration tasks is discussed later on (section 2.3.4).

2.2.3   Results

Results will be reported first for the whole data set pooled across stimulus presentation times and then separately for each subject group. As the immediate apprehension of numerosity was aimed to be studied in this experiment, emphasis was given on short presentation times and more subjects were tested with 50 ms stimulus presentation time ($n = 21$) than with 84 ms ($n = 7$) and 150 ms ($n = 6$). Consequently, at higher presentation times, data are noisier and statistical power is not as high as in the 50 ms group.

Accuracy

Mean accuracy of the data pooled across stimulus presentation times is plotted in Fig. 2.2, whereas the following Fig. 2.3 depict the performance for each stimulus presentation time separately. Likewise, the results of statistical analyses are reported in Table 2.1 for the pooled data set, and then separately for the 50 ms group (Tab. 2.2), the 84 ms group (Tab. 2.3) and the 150 ms group (Tab. 2.4).

Pooled Data Set   As expected, accuracy declined with increasing target number, forming a sigmoidal curve without clear-cut discontinuity between the subitizing and the counting range (Fig. 2.2). In the control condition, performance was close to ceiling at the low numerosities (1 and 2) and at chance at numerosities 8 and 9 (one-sample t-test for above chance performance: $t \leq .72$, $p \geq .47$). In the pre-cue condition, performance reached chance level at numerosity 8 ($t = .57$, $p = .57$) and in the post-cue condition already at numerosities 6 and above ($t \leq 1.7, p \geq .10$). Accuracy was zero at numerosity 9 in both cued conditions. It has to be pointed out that in the cued conditions, trial frequency was very low at numerosities 6 - 9 and statistical power

therefore reduced. Higher numerosities mainly served as catch trials in the cued conditions, so emphasis should be given to the results in the lower numerosity ranges. However, the fact that accuracy in the higher numerosities was low indicates that verbal counting was successfully prevented and subjects were forced to rely on estimation and true subitizing.

There was a highly significant main effect of experimental condition as well as significant differences between each of the conditions (Table 2.1). The pre-cue condition exhibited lower accuracy than the control condition across all numerosities, and crucially, the post-cue condition led to even more decreased accuracy compared to the pre-cue condition. This was true in all numerosity ranges, even in the counting range (5-9) where statistical power was reduced.

As expected from the overall data pattern, there was a significant main effect of target number. Target number interacted with condition, apparent in a slight flattening of curves from control to pre-cue to post-cue. Pre-cue and post-cue conditions were compared separately for this interaction effect. Target number and condition interacted significantly in the whole numerosity set ($F(8,248) = 2.64$, $p = .009$), but did not when subitizing range and counting range were analysed separately (counting range: $F(4,132) = 2.02$, $p > .05$, subitizing range: $F(3,93) = 2.36$, $p > .05$).

Accuracy increased significantly with increasing stimulus presentation times. Post-hoc comparisons revealed that this effect was mainly driven by an accuracy benefit when stimulus exposure was increased from 50 ms to higher presentation times. The increase from 84 ms to 150 ms, however, did not lead to a significant rise in accuracy. The effect of stimulus presentation time interacted with target number, accuracy in the lower numerosities dropped less quickly the longer stimulus exposure was (compare plots in Fig. 2.3). Stimulus presentation time did not interact with condition, indicating that the overall effect of experimental manipulation did not change across subgroups. The significant three-way interaction (condition x stimulus presentation time x target number) therefore must have been mainly driven by the interaction between target number and stimulus presentation time.

50 ms Stimulus Presentation Time   The performance at 50 ms stimulus presentation time mirrored the performance in the pooled data set (Fig. 2.3 and Table 2.2). Accuracy was signif-
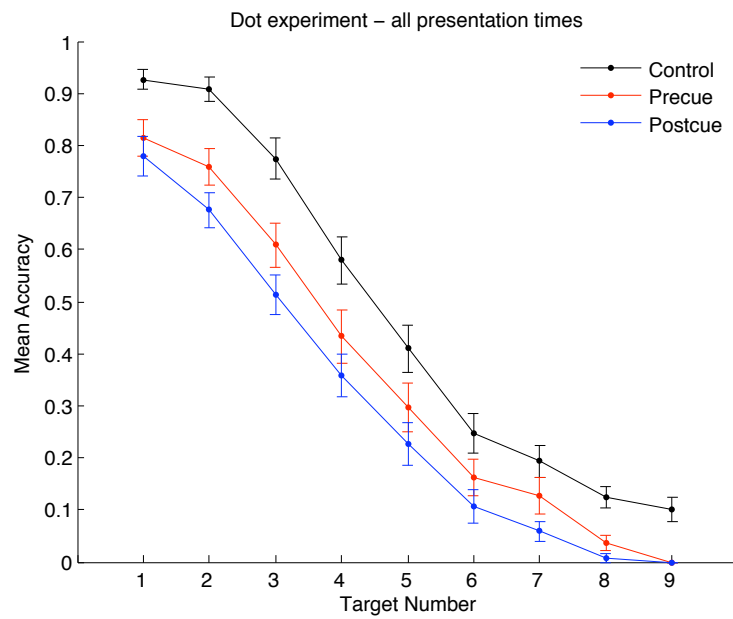
Figure 2.2: Exp. 1.1 Accuracy - All Stimulus Presentation Times Pooled.

Accuracy data (proportion correct) as a function of target number for the data set pooled across all stimulus presentation times. $n = 34$. Note that the amount of trials differ between the control and the cued conditions and across target numbers. Error bars indicate 1 standard error of the mean (SEM).

Exp. 1.1 Accuracy - All Stimulus Presentation Times

		Numerosity Range		
		1 - 9	**1 - 4**	**5 - 9**
Within-Subject Effects				
Condition	p	<.001	<.001	<.001
	F(df)	130.7 (2,62)	61.8 (2,62)	43.8 (2,62)
Target Number	p	<.001	<.001	<.001
	F(df)	279.1 (8,248)	78.1 (3,93)	68.2 (4,124)
Condition x Target Number	p	.011	.003	n. s.
	F(df)	2.01 (16,496)	3.4 (6,186)	1.35 (8,248)
Between-Subject Effects				
Stimulus Presentation Time	p	<.001	<.001	<.001
	F(df)	21.32(2,31)	14.33 (2,31)	23.06 (2,31)
Presentation Time x Condition	p	n. s.	n. s.	n. s.
	F(df)	1.41 (4,62)	2.26 (4,62)	2.2 (4,62)
Presentation Time x Target Number	p	<.001	.001	<.001
	F(df)	5.0 (16,248)	4.43 (6,93)	7.8 (8,124)
Pres. Time x Target Number x Cond.	p	<.001	.039	<.001
	F(df)	2.5 (32,496)	1.9 (12,186)	2.8 (16,248)
Post-hoc Comparisons				
Condition *pre- vs postcue - planned*	p	<.001	<.001	.002
control vs precue	p	<.001	<.001	<.001
control vs postcue	p	<.001	<.001	<.001
Pres. Time *50 ms vs 84 ms*	p	.002	.010	.002
50 ms vs 150 ms	p	<.001	<.001	<.001
84 ms vs 150 ms	p	n. s.	n. s.	n. s.

Table 2.1: Exp. 1.1 Accuracy - All Presentation Times

Repeated Measurement ANOVA results for accuracy data of Exp. 1.1. pooled across stimulus presentation times. $n = 34$ (50 ms: $n = 21$, 84 ms: $n = 7$, 150 ms: $n = 6$). Non-significant effects indicate $p > .05$. p values in post-hoc comparisons are Bonferroni-corrected except where planned.

icantly impaired in the cued conditions compared to the control condition, and particularly in the post-cue condition compared to the pre-cue condition. This was again true in all numerosity ranges.

Condition and target number interacted in the whole numerosity range due to a significant interaction in the counting range. The interaction was not significant in the subitizing range. Again, as the data pattern in the counting range is noisy due to reduced statistical power, interaction effects are not very meaningful.

84 ms and 150 ms Stimulus Presentation Time   Subjects' performance became more accurate with higher stimulus exposure times, particularly in the lower numerosities (see Fig. 2.3). The main effect of experimental condition remained, but was now mainly driven by a difference between pre- and post-cue and control and post-cue conditions (Tables 2.3 and 2.4). Performance in the pre-cue condition did not differ from the control condition anymore. The impairment in the post-cue condition compared to the pre-cue condition was driven by a significant effect in the subitizing range, but did not occur in the counting range. Again, this might be due to reduced statistical power in this range. There was only one interaction effect in the subitizing range with accuracy dropping faster in the cued conditions compared to the control condition. This interaction was not significant in the counting range and overall.

Reaction Times

Reaction times are reported for correct trials only. As performance was at chance for numerosities above 6 (numerosity 8 in the control condition), the further selection of correct trials would have decreased statistical power even more than already due to experimental design, so only reaction times from numerosities 1 to 5 were taken into account. Subitizing and counting range are therefore not considered separately and statistical results are reported for numerosity range 1-5. Statistical tests were carried out also for the subitizing range (1-4) and yielded the same effects unless stated otherwise.
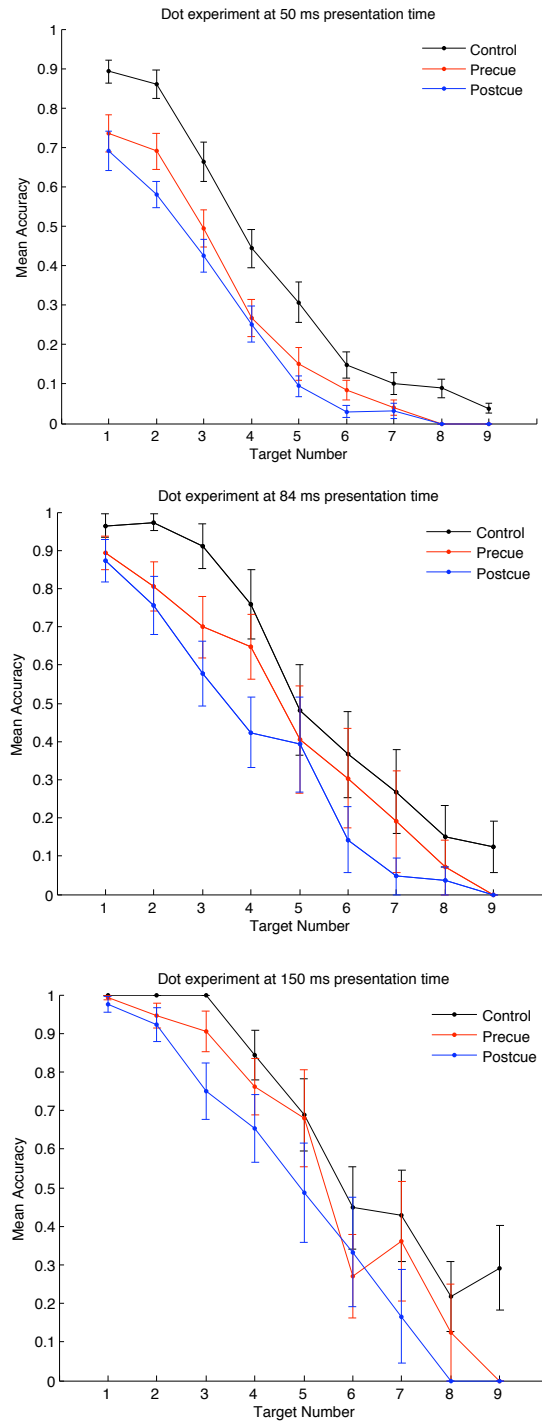
Figure 2.3: Exp. 1.1 Accuracy - All Stimulus Presentation Times Separately.

Accuracy data (proportion correct) as a function of target number plotted separately for 50 ms, 84 ms and 150 ms stimulus presentation times. 50 ms: $n = 21$, 84 ms: $n = 7$, 150 ms: $n = 6$. Error bars reflect 1 SEM.

Exp. 1.1 Accuracy - 50 ms Presentation Time

		Numerosity Range		
		1 - 9	**1 - 4**	**5 - 9**
Within-Subject Effects				
Condition	p	<.001	<.001	<.001
	F(df)	120.5 (2,40)	71.5 (2,40)	21.4 (2,40)
Target Number	p	<.001	<.001	<.001
	F(df)	207.5 (8,160)	113.8 (3,60)	22.01 (4,80)
Condition x Target Number	p	<.001	n. s.	<.001
	F(df)	4.21 (16,320)	.473 (6,120)	5.39 (8,160)
Post-hoc Comparisons				
Condition precue vs postcue - planned	p	<.001	<.001	.009
control vs precue	p	<.001	<.001	<.001
control vs postcue	p	<.001	<.001	.027

Table 2.2: Exp. 1.1 Accuracy - 50 ms Presentation Time

Repeated Measurement ANOVA results for accuracy data of Exp. 1.1. for 50 ms stimulus presentation time. $n = 21$. Non-significant effects indicate $p > .05$. p values in post-hoc comparisons are Bonferroni-corrected except where planned.

Exp. 1.1 Accuracy - 84 ms Presentation Time

		Numerosity Range		
		1 - 9	**1 - 4**	**5 - 9**
Within-Subject Effects				
Condition	*p*	<.001	<.001	.002
	F(df)	28.12 (2,12)	17.17 (2,12)	10.47 (2,12)
Target Number	*p*	<.001	<.001	<.001
	F(df)	100.7 (8,48)	21.12 (3,18)	18.00 (4,24)
Condition x Target Number	*p*	n. s.	.001	n. s.
	F(df)	1.68 (16,96)	4.68 (6,36)	.815 (8,48)
Post-hoc Comparisons				
Condition precue vs postcue - planned	*p*	.038	.024	n. s.
control vs precue	*p*	.020	n. s.	n. s.
control vs postcue	*p*	<.001	.001	.003

Table 2.3: Exp. 1.1 Accuracy - 84 ms Presentation Time

Repeated Measurement ANOVA results for accuracy data of Exp. 1.1. for 84 ms stimulus presentation time. $n = 7$. Non-significant effects indicate $p > .05$. p values in post-hoc comparisons are Bonferroni-corrected except where planned.

Exp. 1.1 Accuracy - 150 ms Presentation Time

		Numerosity Range		
		1 - 9	**1 - 4**	**5 - 9**
Within-Subject Effects				
Condition	p	<.001	.003	.002
	F(df)	24.72 (2,10)	10.30 (2,10)	11.62 (2,10)
Target Number	p	<.001	<.001	<.001
	F(df)	55.43 (8,40)	12.12 (3,15)	14.30 (4,20)
Condition x Target Number	p	n. s.	.039	n. s.
	F(df)	1.37 (16,80)	2.58 (6,30)	1.20 (8,40)
Post-hoc Comparisons				
Condition precue vs postcue - planned	p	.016	.022	n. s.
control vs precue	p	n. s.	n. s.	n. s.
control vs postcue	p	.001	.053	.005

Table 2.4: Exp. 1.1 Accuracy - 150 ms Presentation Time

Repeated Measurement ANOVA results for accuracy data of Exp. 1.1. for 150 ms stimulus presentation time. $n = 6$. Non-significant effects indicate $p > .05$. p values in post-hoc comparisons are Bonferroni-corrected except where planned.

Pooled Data Set   As expected, reaction time increased with target number, again without a clear-cut discontinuity in the curve (Fig. 2.4). There was a strong effect of experimental condition, driven by a significant increase in reaction time in the post-cue condition compared to the pre-cue and the control condition (Table 2.5). As opposed to accuracy data, however, reaction times did not differ in the pre-cue and the control condition. As expected, there was a clear effect of target number. No interaction between target number and condition occurred, indicating that the increase in reaction time in the post-cue condition was equally strong in all numerosities.

Stimulus presentation time did not have an effect on reaction times, subjects responded neither slower nor faster with increasing stimulus exposure. There was neither an interaction effect with condition, nor target number, but a significant three-way interaction. However, when considering the subitizing range separately (1-4), this three-way interaction disappeared ($F(12,186) = .48$, $p > .05$).

50 ms, 84 ms and 150 ms Stimulus Presentation Time   As there was no main effect of stimulus presentation time, performance in the different subject groups mirrored the pooled data set. Statistical results for each stimulus presentation time are summarised in Table 2.6. The main effect of condition replicated in all subject groups, particularly the increase in reaction time in the post-cue condition compared to the pre-cue condition. The difference between control and post-cue condition, however, was only significant in the 50 ms group, but not in the 84 ms and the 150 ms group (with exception to the subitizing range (1-4) in the 84 ms group ($p = .006$)). Variance was larger in the latter two groups due to small subject numbers, therefore the effects are not very clear. There was again no interaction between target number and condition, mirroring the constant slowing of responses in the post-cue condition across all numerosities as seen in the pooled data set.

Mean Responses, Response Variance and Weber Fraction

As accuracy data is only informative in a binary manner (it tells whether subjects hit the exact numerosity or not), mean responses and their variance were analysed as a measure of response
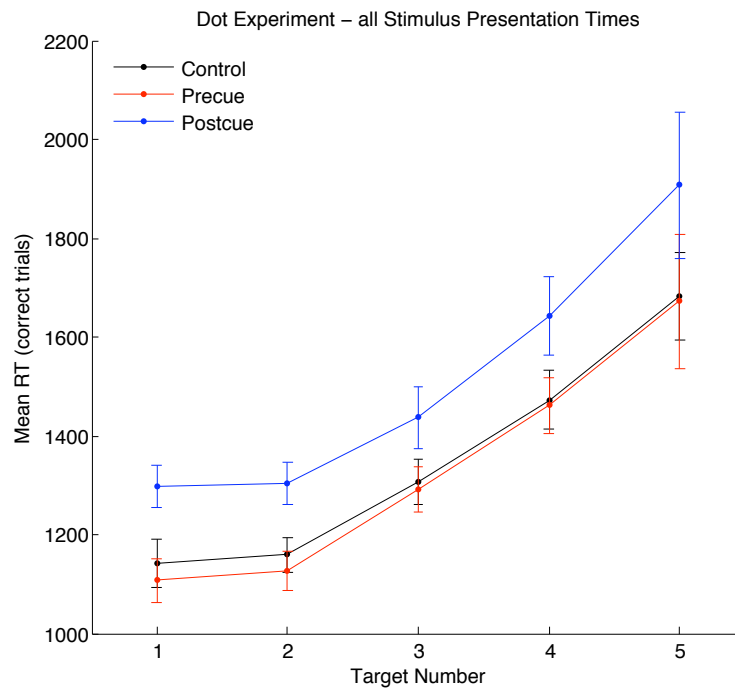
Figure 2.4: Exp. 1.1 Reaction Times - All Stimulus Presentation Times Pooled.

Reaction time data (correct trials only, in ms) as a function of target number for the data set pooled across all stimulus presentation times. $n = 34$. Trial frequency at numerosities above 5 was very low and data not reliable. Error bars reflect 1 SEM.

Exp. 1.1 Reaction Times - All Stimulus Presentation Times

	Numerosity Range 1 - 5	
	F(df)	p
Within-Subject Effects		
Condition	28.5 (2,62)	<.001
Target Number	39.9 (4,124)	<.001
Condition x Target Number	.033 (8,248)	n. s.
Between-Subject Effects		
Stimulus Presentation Time	.57 (2,31)	n. s.
Presentation Time x Condition	.07 (4,62)	n. s.
Presentation Time x Target Number	.85 (8,124)	n. s.
Pres. Time x Target Number x Condition	2.08 (16,248)	.010
Post-hoc Comparisons		
Condition precue vs postcue - planned		<.001
control vs precue		n. s.
control vs postcue		<.001
Pres. Time 50 ms vs 84 ms		n. s.
50 ms vs 150 ms		n. s.
84 ms vs 150 ms		n. s.

Table 2.5: Exp. 1.1 Reaction Times - All Presentation Times

Repeated Measurement ANOVA results for reaction time data of Exp. 1.1. pooled across stimulus presentation times. $n = 34$ (50 ms: $n = 21$, 84 ms: $n= 7$, 150 ms: $n = 6$). Non-significant effects indicate $p >.05$. p values in post-hoc comparisons are Bonferroni-corrected except where planned.

Exp. 1.1 Reaction Times - 50, 84 and 150 ms Presentation Times

		Presentation Time		
		50 ms	**84 ms**	**150 ms**
Within-Subject Effects				
Condition	p	<.001	.005	.014
	F(df)	22.88 (2,40)	8.69 (2,12)	6.74 (2,10)
Target Number	p	<.001	<.001	<.001
	F(df)	39.71 (4,60)	11.78 (4,24)	12.82 (4,20)
Condition x Target Number	p	.002	n. s.	n. s.
	F(df)	3.25 (8,160)	.26 (8,48)	1.62 (8,40)
Post-hoc Comparisons				
Condition precue vs postcue - planned	p	<.001	.007	.001
control vs precue	p	n. s.	n. s.	n. s.
control vs postcue	p	<.001	n. s.	n. s.

Table 2.6: Exp. 1.1 Reaction Times - 50, 84 and 150 ms Presentation Time

Repeated Measurement ANOVA results for reaction time data of Exp. 1.1. for 50, 84 and 150 ms stimulus presentation time separately. All results are for numerosity range 1-5. 50 ms: $n = 21$, 84 ms: $n = 7$, 150 ms: $n = 6$. Non-significant effects indicate $p > .05$. p values in post-hoc comparisons are Bonferroni-corrected except where planned.

deviation. Mean responses inform about the extent of under- and overestimation, whereas mean response variance reflects how widely the responses varied at a given numerosity. As a measure of discriminability, a correlation of Weber fraction was adopted and defined as the response standard deviation divided by the target number. This approach was inspired by Ross (2003) who also applied a correlation of Weber fraction to his data on numerosity discrimination (however, he derived Weber fraction from 90% increment thresholds). It has to be noted that this is not the classical definition of Weber fraction as this paradigm did not comprise the discrimination (or comparison) between two sets of numerosity, but the discrimination of one numerosity from 8 possible alternatives. The employed definition was hypothesised to be closest to the original sense of Weber fraction as it should reflect discrimination variance normalised by the number of targets.

Mean Responses   Mean responses were plotted as a function of target number (pooled data set: Fig. 2.5, subgroups: Fig. 2.6). The dotted black diagonal indicates perfect performance. Data points above the diagonal represent overestimation, whereas data points below the diagonal represent underestimation. Statistical results are shown in Table 2.7 for the pooled data set and in Table 2.8 summarised for each subgroup.

In the pooled data set, subjects on average underestimated from numerosity 4 onwards (One-sample t-tests from perfect performance: t $\leq$ -2.69, $p \leq$ .01), and underestimation increased with increasing target number (repeated measures ANOVA on response deviations F(8,248) = 154.6, $p <$ .001). At numerosity 1, overestimation occurred (t > 2.75, $p <$ .010), whereas performance did not significantly depart from the correct response at numerosities 2 and 3 (t > 1.9, $p <$ .066, except post-cue condition at numerosity 3: t = 3.32, $p =$ .002). However, as accuracy data showed that performance was not 100%, this just indicates that the extent of over- and underestimation was balanced.

There was a main effect of condition in the pooled data set, with significant differences between each condition as revealed by post-hoc tests (Table 2.7). From numerosity 4 onwards, the post-cue condition led to the highest extent of underestimation, whereas subjects underestimated less in the pre-cue condition and least in the control condition. This was particularly

evident in the higher numerosities. Naturally, there was an effect of target number that interacted significantly with condition, confirming the increasing extent of underestimation from control to pre-cue to post-cue condition.

Stimulus presentation time had a significant effect, driven by the difference between the 50 ms group and both the 84 ms and 150 ms group as the post-hoc comparisons indicate. The 84 ms and 150 ms group did not differ significantly from each other. The graphs show that underestimation occurred earlier for 50 ms presentation time (at numerosity 3 in pre- and post-cue conditions, t < 2.77, p < .012) compared to longer presentation times, overall apparent in a significant interaction between target number and stimulus presentation time. Significant underestimation started in the 84 ms group only for the post-cue condition at numerosity 4 (t = -4.28, p = .005), whereas in the control condition, underestimation was apparent from numerosity 7 onwards (t = -5.02 , p = .002). At 150 ms stimulus presentation time, subjects underestimated in the post-cue condition from numerosity 6 onwards (t = -7.0, p = .001), but only from numerosity 8 onwards in the control condition (t = -3.07 , p = .028).

Overestimation occurred in the 50ms group at numerosity 1 in all conditions (t > 2.63, p < .016), whereas it occurred at numerosities 1 and 2 only for the post-cue condition in the 84 ms group (t > 2.53, p < .045). At 150ms subjects overestimated only in the pre-cue condition at numerosity 3 (t = 2.93, p = .033).

Stimulus presentation time did not interact significantly with condition, indicating that prolonged stimulus exposure did not affect the effect of experimental manipulation. There was also no three-way interaction (stimulus presentation time x target number x condition).

As shown in Table 2.7, all effects of condition and their post-hoc comparisons replicated as significant when analysing the different subgroups separately (except for a non-significant difference between control and pre-cue condition at 84ms). Also the effect of target number and its interaction with condition was significant in all subgroups.

Mean Response Variance   The variance of mean responses is plotted in Fig. 2.7. Response variance increased with target number, evident as significant main effect of target number (re-
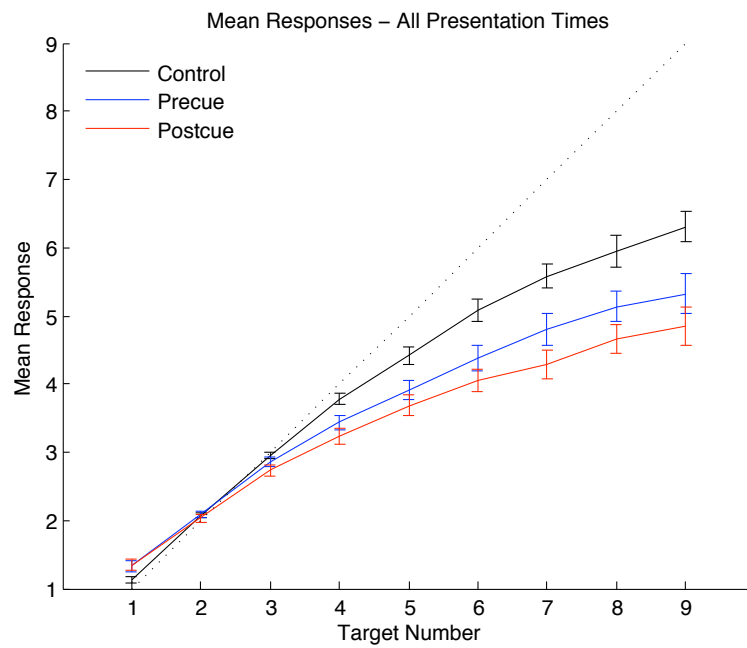
Figure 2.5: Exp. 1.1 Mean Response - All Stimulus Presentation Times Pooled.

Mean response as a function of target number for the data set pooled across all stimulus presentation times. $n = 34$. The dotted line indicates perfect performance. Error bars reflect 1 SEM.

Exp. 1.1 Mean Responses - All Stimulus Presentation Times

	Numerosity Range 1 - 9	
	F(df)	p
Within-Subject Effects		
Condition	41.7 (2,62)	<.001
Target Number	368.2 (8,248)	<.001
Condition x Target Number	15.8 (16,496)	<.001
Between-Subject Effects		
Stimulus Presentation Time	15.8 (2,31)	<.001
Presentation Time x Condition	.86 (4,62)	n. s.
Presentation Time x Target Number	13.9 (16,248)	<.001
Pres. Time x Target Number x Condition	.89 (32,496)	n. s.
Post-hoc Comparisons		
Condition precue vs postcue - planned		<.001
control vs precue		<.001
control vs postcue		<.001
Pres. Time 50 ms vs 84 ms		.001
50 ms vs 150 ms		<.001
84 ms vs 150 ms		n. s.

Table 2.7: Exp. 1.1 Mean Responses - All Presentation Times

Repeated Measurement ANOVA results for mean response data of Exp. 1.1. pooled across stimulus presentation times. $n = 34$ (50 ms: $n = 21$, 84 ms: $n = 7$, 150 ms: $n = 6$). Non-significant effects indicate $p > .05$. p values in post-hoc comparisons are Bonferroni-corrected except where planned.
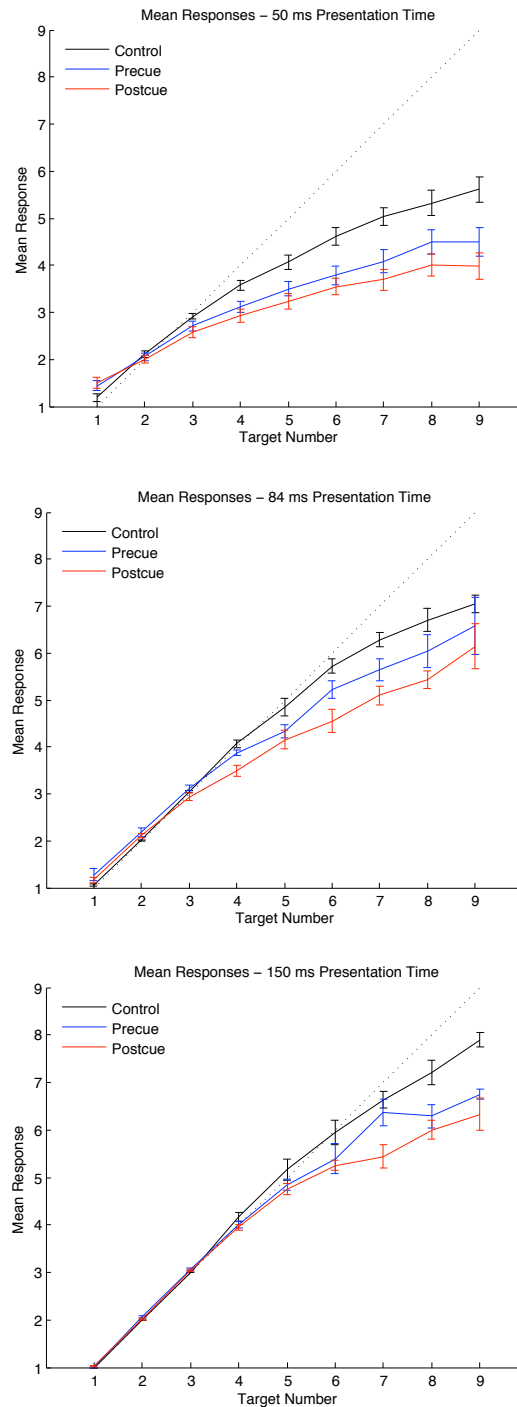
Figure 2.6: Exp. 1.1 Mean Response - All Stimulus Presentation Times Separately.

Mean response as a function of target number for 50 ms, 84 ms and 150 ms stimulus presentation times separately. 50 ms: $n = 21$, 84 ms: $n = 7$, 150 ms: $n = 6$. The dotted line indicates perfect performance.

Exp. 1.1 Mean Responses - 50, 84 and 150 ms Presentation Times

		Presentation Time		
		50 ms	**84 ms**	**150 ms**
Within-Subject Effects				
Condition	*p*	<.001	.005	<.001
	F(df)	40.25 (2,40)	12.48 (2,12)	25.91 (2,10)
Target Number	*p*	<.001	<.001	<.001
	F(df)	113.5 (8,160)	165.8 (8,48)	439.9 (8,40)
Condition x Target Number	*p*	<.001	<.001	<.001
	F(df)	15.6 (16,320)	3.34 (16,96)	7.21 (16,80)
Post-hoc Comparisons				
Condition precue vs postcue - planned	*p*	.003	.018	.034
control vs precue	*p*	<.001	n. s.	.025
control vs postcue	*p*	<.001	.006	.003

Table 2.8: Exp. 1.1 Mean Responses - 50, 84 and 150 ms Presentation Time

Repeated Measurement ANOVA results for mean response data of Exp. 1.1. for 50, 84 and 150 ms stimulus presentation time separately. All results are for numerosity range 1-9. 50 ms: $n = 21$, 84 ms: $n = 7$, 150 ms: $n = 6$. Non-significant effects indicate $p > .05$. p values in post-hoc comparisons are Bonferroni-corrected except where planned.
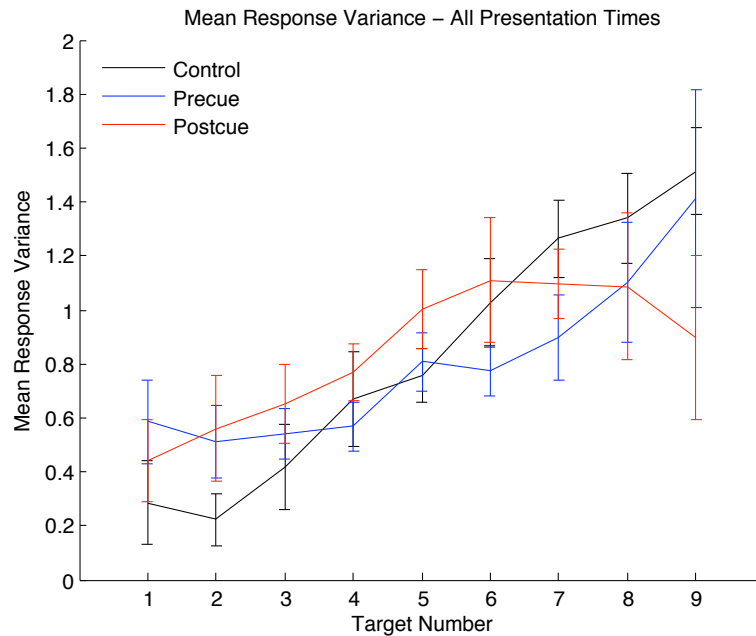
Figure 2.7: Exp. 1.1 Mean Response Variance - All Stimulus Presentation Times Pooled.

Mean response variance as a function of target number for the pooled data set. $n = 34$. Error bars reflect 1 SEM.

peated measurement ANOVA: $F(8,248) = 10.95$, $p < .001$). However, there were no other significant effects: variances did neither differ across experimental conditions ($F(2,62) = .71$, $p > .05$) nor across stimulus presentation times ($F(2,31) = 1.43$, $p > .05$). None of the factors interacted with each other.

Weber Fraction   Weber fractions were calculated as a measure of discriminability. The higher the Weber fraction, the lower the discriminability (i.e. following the original definition, the more difference between stimulus size is needed for successful discrimination). Weber fractions averaged across all stimulus presentation times are plotted in Fig. 2.8.

Weber fractions were consistently higher in the cued conditions than in the control
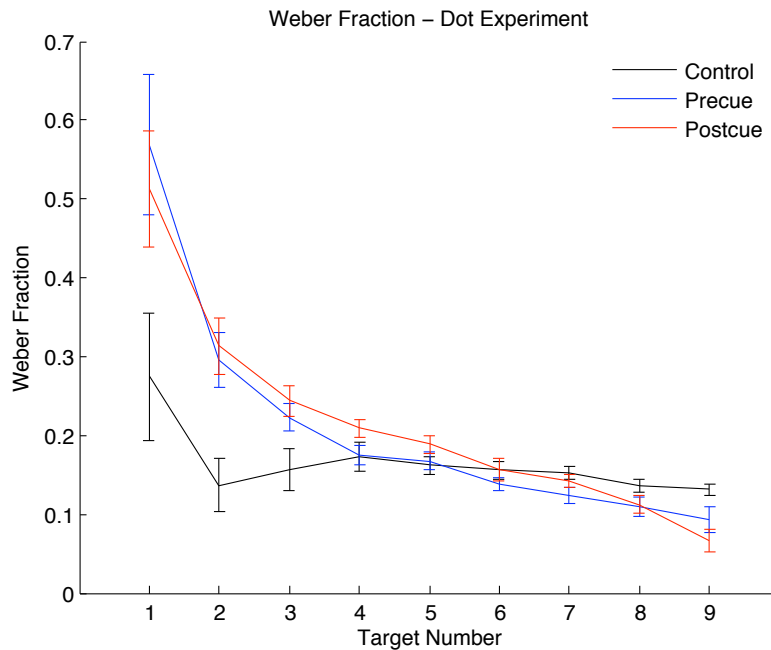
Figure 2.8: Exp. 1.1 Weber Fraction - All Stimulus Presentation Times Pooled.

Weber fraction (response standard deviation / target number) as a function of target number for the pooled data set. $n = 34$. Error bars reflect 1 SEM.

condition, particularly in the subitizing range. There was a main effect of condition ($F(2,66)$ = 13.29, $p < .001$) due to the difference between the control and both cued conditions (control versus pre-cue: $p = .003$, control versus post-cue: $p < .001$). However, there was no difference between the pre- and post-cue condition,($p > .05$), except in the counting range when analysed separately ($p = .007$). Weber fractions decreased with target number ($F(8,264) = 21.58$, p $<.001$), and the effect was replicated also when analysing all three experimental conditions separately. The effect of target number interacted with condition ($F(16,528) = 10.97$, $p < .001$). All effects replicated when analysing subitizing and counting range separately.

Control Condition: Effects of Stimulus Heterogeneity

As mentioned in the introduction (section 2.2.1), the control condition was designed so as to study the effects of luminance heterogeneity on enumeration. In this analysis, the trials of the control condition were divided into trials with homogeneous / uni-coloured stimuli (only white or only black dots) and heterogeneous / mixed-coloured stimuli (black and white dots). Fig. 2.9 shows the accuracy and reaction time data for both uni-coloured and mixed-coloured trials for the pooled data set.

There was no significant difference between homogeneous and heterogeneous trials across the whole range of target numbers, neither for accuracy ($F(1,31) = .635$, $p = .43$) nor for reaction times ($F(1,31) = 1.09$, $p = .74$). As found already in the analysis including all experimental conditions, there was a significant effect of target number (Accuracy: $F(7,217) = 132.5$, $p < .001$, RT: $F(7,217) = 19.01$, $p < .001$) and accuracy differed across stimulus presentation times ($F(2,31) = 18.2$, $p < .001$) whereas reaction times did not ($F(2,31) = .461$, $p = .63$).

2.2.4   Discussion

This experiment investigated the effect of attentional set on visual enumeration. Subjects viewed a briefly presented display of black and white dots and enumerated either all items (control condition) or either the white or the black subset, depending on a cue received either before (pre-cue condition) or after (post-cue condition) stimulus presentation. It was hypothesised that if small numerosity information is retrieved in a pre-attentive and parallel manner, it should not matter whether subjects know in advance to which subset to direct their attention to. Hence, enumeration performance should be equally accurate and fast in the pre-cue and the post-cue condition in the small numerosities.

Effect of Experimental Conditions

The significant effects of experimental conditions reject the above hypothesis: both enumeration accuracy and reaction times differed significantly across the whole range of numerosities, with
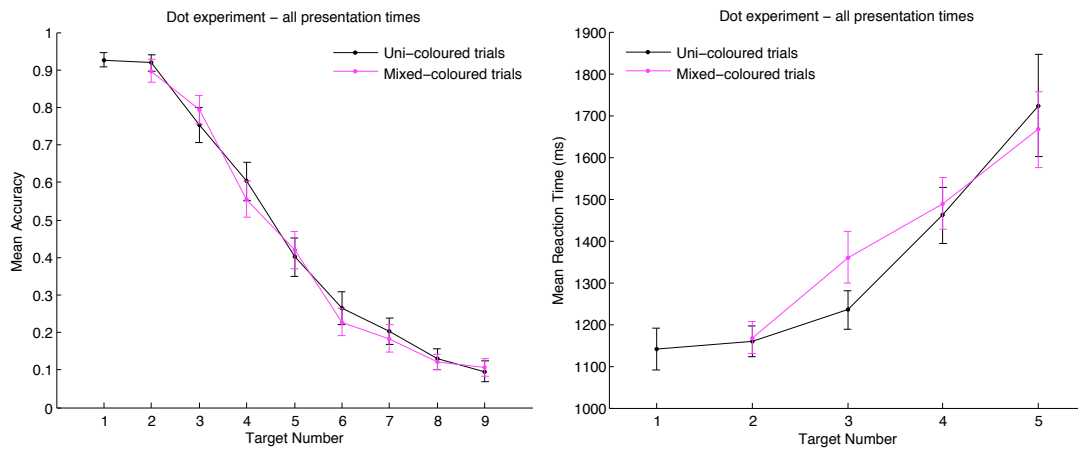
Figure 2.9: Exp. 1.1 Control Condition - Results by Stimulus Heterogeneity.

Accuracy (left) and reaction time data (right) as a function of target number for the data of the control condition (pooled across all stimulus presentation times) separately for uni-coloured (homogeneous) and mixed-coloured (heterogeneous) trials. n = 34. Error bars reflect 1 standard error of the mean (SEM).

the post-cue condition resulting in consistently lower accuracy and higher reaction times than the pre-cue condition. That is, not being able to attend to the relevant subset results in a behavioural disadvantage compared to when attention can be drawn to the subset to report. Therefore, the correct numerosity information does not appear to be automatically retrieved from two subsets in parallel, at least not to the extent that it is available for conscious report. Reduced accuracy in the post-cue condition compared to the pre-cue condition was particularly apparent in the subitizing range in all subgroups, independent of stimulus presentation times, indicating that subitizing performance was systematically affected by the employed experimental manipulation. This argues against the notion of subitizing as an automatic, parallel or pre-attentive process. The absence of a significant interaction between condition and target number in the subitizing range demonstrates that the effect was constant across all 4 numerosities. Likewise, accuracy dropped also in the counting range given sufficient statistical power. Given that estimating higher numerosities is a more difficult and more resource demanding process, particularly with the employed short presentation times, this might not be a surprising result. What might appear surprising, though, is the fact that the absolute accuracy difference between post-cue and pre-cue conditions was almost constant throughout the whole numerosity range. However, the differences in accuracy might better be considered differentially with respect to the accuracy level of each target number. For example, an accuracy decrease from 76% to 68% at numerosity 2 is a relative drop of 11% and therefore a less strong effect than an accuracy drop from 16% to 11% at numerosity 6 with a relative decrease of 31%. When considering these relative accuracy differences, the effect of post-cueing versus pre-cueing increased with numerosity. Supposing that higher numerosity judgement involves higher perceptual load, this is an expected finding.

An effect of cueing was also found in reaction time data: reaction times were consistently higher in the post-cue condition compared to the pre-cue condition. This slowing down was constant across all numerosities, apparent in the absence of an interaction effect of condition with target number. This confirms the conclusions made from accuracy data that post-cueing leads to a behavioural disadvantage compared to pre-cueing. However, reaction time data need to be considered with care in this paradigm. First, accuracy rather than speed was emphasised to subjects. Second, reaction times in the post-cue condition might have included the processing of the cue (reaction times were measured from stimulus onset until response, in the post-cue

condition corrected for the duration of the mask). Accuracy is therefore a more meaningful measure in this paradigm (as it typically is in paradigms with short stimulus exposure). However, the least that can be inferred from reaction time data is that the accuracy effect was not due to a speed-accuracy trade-off.

Accuracy is a binary measure and gives an insufficient picture on enumeration performance as it does not take the amount of response deviation into account. For example, if a subject responded 4 to a target number of 5, then the judgement was relatively correct, but counted as a miss in accuracy data. As for the other extreme, if a subject responded 1 to target number 9, then this was a much more severe misjudgement which was not captured in binary accuracy measurement. Therefore, mean responses and their variances were analysed as a measure of the extent, direction and variability of response deviation. That is, accuracy reflects the proportion of deviant trials, whereas mean responses inform about the average extent and direction of deviation and response variances about the variability of responses. Furthermore, the notion of Weber fraction was adopted as a measure of discriminability.

In general, mean responses revealed underestimation from numerosity 4 onwards and overestimation at numerosity 1. At numerosities 2 and 3, over- and underestimation seemed to be balanced, as responses on average did not depart from the correct response number (as accuracy dropped between conditions, however, there was deviation from the correct response). Given that numerosity 1 cannot be underestimated, overestimation may just reflect deviation from the correct response. Accordingly, overestimation is impossible at numerosity 9, therefore underestimation is a more likely finding in the higher numerosities. In general, underestimation of higher numerosities is a common finding in random dot patterns (e.g. Ginsburg, 1991; Mandler & Shebo, 1982). Furthermore, a probabilistic argument can be made about underestimation. When the brain receives a certain amount of signals from the environment, it is more likely that some signals may not reach the threshold to be detected and processed, resulting in fewer signals being processed, rather than the brain "making up" signals that are not actually there. Underestimation in itself may therefore not be a surprising finding.

With respect to the employed experimental manipulation, post-cueing led to more underestimation from numerosity 4 onwards than pre-cueing. This could mean that post-cueing

either results in noisier response behaviour in general or in a systematic underestimation of the target subset. The results from response variances suggest a systematic effect: response variances did not differ between experimental conditions, indicating that cueing did not result in a wider spreading of responses, i.e. more random response behaviour. This implies that subjects' variability in deviation from the actual target number was more or less equal across conditions. It appears that post-cueing at larger numerosities led to a systematic omission of items at report. With respect to subitizing, subjects deviated from the correct response to a similar extent in both directions up to numerosity 3. The decrease in accuracy in the subitizing range did therefore not arise from systematic under- or overestimation.

Weber fraction was more or less constant in the control condition (though noisy in the subitizing range, leading to an effect of target number) as would be expected from studies by Ross (2003), for example. Discriminability was considerably lower in the cued conditions, particularly in the subitizing range. However, there was no effect of pre- versus post-cueing, suggesting that the decrease in discriminability was not caused by attentional set, but by the difference between control and cued conditions as discussed in the following.

With regard to the control condition, accuracy was consistently higher than in the cued conditions. Two factors might have contributed to this finding. First, averaging trials by target number entails a crucial difference between the control and the cued conditions. In the control condition, the number of items to enumerate was equal to the total number of items present in the display. In the cued conditions, however, the data is collapsed over all possible combinations of distractor numerosities given a certain target. For example, at target number 3 there were just 3 items present in the control condition, but up to 10 in the cued condition. This leads to a facilitation in the control condition compared to the cued conditions, as the overall amount of information can be used a hint for a more accurate numerosity estimation and the absence of distractors prevents noise. In the cued conditions however, the number of distractors and the total amount of visual information varied greatly, leading to increased perceptual load, greater noise and thus higher error rates. Second, target-distractor segregation in the cued conditions might have contributed to the observed accuracy decrease. There was no effect on reaction times, however. This suggests that the judgement of the target numerosities became more erroneous (i.e. noisier) with the presence of distractors, but that target-distractor discrimination itself did

not render the involved processing more time consuming.

In fact, the interpretation that increased visual noise is responsible for the difference between the control and the cued conditions is confirmed by the pattern of Weber fraction. Discriminability was lower in both cued conditions compared to the control condition, suggesting that the presence of distractors, and possibly the effort of segregating them from targets, limited discriminability. Attentional set, however, did not interfere with discrimination ability, thus the effect of cueing cannot be attributed to an effect of visual noise. Whether target-distractor discrimination was affected by cueing, however, is addressed in Experiment 1.3.

Effect of Stimulus Presentation Times

Stimulus presentation time was varied between 50 ms, 84 ms and 150 ms. 50 ms was regarded as the shortest presentation time that ensured conscious viewing and 150 ms as the longest presentation time avoiding saccades.

In general, prolonged stimulus exposure led to higher accuracy, but did not affect reaction times. The accuracy benefit was particularly evident when stimulus presentation was increased from 50 ms to either 84 ms or 150 ms, but did not occur between 84 and 150 ms (possibly due to lacking statistical power). Mean response data mirrored this pattern: underestimation started at higher numerosities in the higher stimulus presentation times compared to the 50 ms group. This suggests that longer stimulus exposure gave subjects more time to inspect the display and possibly process and retain the numerosity information more correctly, resulting in less noisy response behaviour and fewer omissions of items.

Importantly, however, stimulus presentation time did not affect the effect of experimental manipulation. The decrease in accuracy between the pre-and post-cue condition (particularly in the subitizing range) as much as the increasing underestimation in the higher numerosities was consistent in all subgroups. This shows that, although numerosity judgement overall became better with prolonged stimulus exposure, post-cueing still had a more detrimental effect on performance than pre-cueing. Consequently, being able to view the stimulus longer did not ameliorate the attentional effect of post-cueing.

In the 150 ms group, there was no accuracy difference between pre-cue and control condition anymore (mirroring the reaction time data), possibly indicating that the cost of target-distractor segregation or the effect of stimulus noise disappears with sufficient stimulus exposure.

Effect of Stimulus Heterogeneity

The control condition was designed such that the effect of stimulus heterogeneity on enumeration could be studied. The question at hand was whether enumeration performance changes when subjects judged the numerosity of either only black or only white dots compared to both black and white dots. The absence of any effect of luminance heterogeneity shows that a mixed-coloured display did neither help subjects to enumerate more correctly (using a subitize-and-add strategy) nor hinder subjects to enumerate correctly by having to ignore task-irrelevant perceptual information.

This result confirms a recent study by Watson and Maylor (2006) who also failed to find a difference in subitizing reaction times when using displays containing either uni-coloured displays (red or green dots) or mixed-coloured displays (red and green dots). The same negative result was found even when presenting red, green or mixed displays in a blocked fashion so that subject were able to "set" themselves to a specific colour, thus avoiding redistribution of attentional weights from one colour map to another between trials. Watson and Maylor (2006) explain their results in light of the visual search literature and argue that targets that vary along a single feature dimension (e.g. colour) can be detected (and discriminated from distractors) without RT cost (Müller, Heller & Ziegler, 1995). This suggests that subsets differing along a single dimension should not invoke confounding perceptual information that hinders detection and subsequent enumeration. Subsets differing across more than one dimension, however, might. In fact, Found and Müller (1996) demonstrated that discrimination between 3 and 4 targets is slower when targets vary across dimensions (colour and orientation) compared to when targets vary within dimensions (just colour or orientation).

It has to be noted that Found and Müller (1996) found a RT cost for heterogeneously coloured trials compared to homogeneously coloured trials (but not for hetero - or homogeneously oriented stimuli), as much as a study by Puts and de Weert (1997) did. However, both studies

employed a 2-alternative forced choice (2AFC) task instead of asking subjects to report the exact numerosity. That means, subjects had to make a more/less judgement rather than a true numerosity judgement. Such a comparative judgement could have encouraged subjects to use the global properties of the display such as overall luminance or "redness". This sort of judgment might be more efficient in single-colour conditions than in mixed-colour conditions (Watson & Maylor, 2006).

When studying approximation of larger numerosities in many spatially overlapping and differently coloured subsets, also Halberda et al. (2006) found that the number of colours the stimulus was composed of affected neither enumeration error rates nor speed, even when there were up to 6 different colour subsets present.

Most of the studies discussed here only anticipated a *cost* in enumeration with heterogeneous displays, but only few considered the possibility of an enumeration *benefit*. When 3 black and 2 white items are displayed and subitizing occurs truly in parallel, then a subitize-and-add strategy should result in a more accurate judgement, particularly when the sum of both subsets exceeds the subitizing range. There might not be a reaction time benefit as the addition operation might slow down responses, but an accuracy benefit could be expected at higher numerosities that can be broken down in two subitizable sets. Indeed, Atkinson, Francis & Campbell (1976) report an accuracy benefit when linearly arranged dots can be grouped into two subsets by either a kink in the linear arrangement (with a sufficiently large angle) or by a sufficiently large space. For such divided dot arrangements, subjects were error-free in enumerating up to 8 items, though no reaction time benefit was observed. When dots could be grouped into two subsets by either colour, smaller array angles or smaller spacing, enumeration of 5 to 9 items was not error-free, but an accuracy benefit (though no RT benefit) was observed compared to ungrouped arrangements. The authors propose that only in the former case subitizing occurred in both subsets independently but not in the latter case. Given that no other study found an accuracy benefit for subsets separated by colour it appears that such grouping cue does not induce automatic and parallel subitizing. Whether the grouping by spatial arrangement induce real parallel subitizing (as, for example, also suggested by Wender & Rothkegel, 2000) or whether it just facilitates judgement by correlating parameters such as area covered or overall luminance, remains to be investigated.

Possible Criticisms

The results show that enumeration, with subitizing as a special focus, is affected by subjects' advance knowledge of which of two subsets to enumerate. In accordance with the originally made hypothesis, this argues against an automatic, pre-attentive and parallel subitizing mechanism. However, this conclusion assumes that cueing presents an attentional manipulation specifically directed at enumeration and several criticisms can be made of this interpretation.

First, it could be argued that the attentional manipulation employed might not have interfered with enumeration per se, but with stimulus processing steps on which enumeration is contingent. For example, enumeration presumably involves detection and individuation of items, segregation from distractors and grouping of targets into subsets to enumerate. Any of those processes could be attention-demanding under the conditions of the current paradigm, and/or affected by the cueing manipulation, whereas the single enumeration process might actually occur pre-attentively. The following experiments 1.2., 1.3. and 1.4. were designed to address this issue, investigating how much of the observed effect of post-cueing is due to item detection, discrimination and perceptual grouping.

Second, a further criticism might be that the cueing manipulations are not entirely attentional manipulations but also those of memory. In the pre-cue condition, subjects needed to encode and hold the numerosity of only one subset in working memory, whereas in the post-cue condition, subjects had to process and retain the numerosity of two subsets in order to give an equally accurate response. Therefore, the post-cue condition could have imposed a higher memory load than the pre-cue condition, leading to lower accuracy and longer response times. Possible arguments against such an interpretation will be discussed in detail in the general discussion of this chapter (section 2.6.2).

The results presented here might seem in contradiction with the results of Halberda et al. (2006) who also employed a cueing paradigm to study approximate enumeration. They presented up to 35 dots in up to 6 differently coloured, spatially intermixed subsets and investigated how many *subsets* can be enumerated in parallel. Also here, pre- and post-cueing was used as an attentional manipulation. Halberda et al. (2006) found no difference in error rates between

pre- and post-cueing when two colour subsets are present, but subjects made significantly more errors in the post-cue condition with more than 3 subsets to enumerate. The authors interpreted the first occurrence of a significant cueing effect at 3 subsets as criterion to conclude that 3 sets can be enumerated in parallel (although the data suggest that 2 subsets were in fact the true limit). There are however, a few differences between the current paradigm and that of Halberda et al. (2006). First, they report results for numerosities above 8 and studied approximation. Second, subjects were able to view the stimulus for 500 ms - a sufficiently long time to scan through the display and approximate each of the two subsets one after each other. The fact that error rates increased in the post-cue condition compared to the pre-cue condition when there were more than 3 subsets present suggests that the time of stimulus presentation might not have been sufficient to approximate all subsets serially. Thus, the parallel enumeration that Halberda et al. claim might have been in fact serial. In the current paradigm, this problem was prevented by using much shorter stimulus presentation times that did not allow eye movements or scanning through the display. Results therefore suggest that when subjects have to rely on true parallel numerosity perception, two subsets cannot be enumerated in parallel.

2.3   Experiment 1.2 Effects of Perceptual Grouping

2.3.1   Rationale

The results of Experiment 1.1 left it unclear whether cueing affected the enumeration process per se or whether it affected perceptual grouping of both subsets. In order to address this issue, the stimuli were modulated such that the grouping criterion was perceived depth rather than luminance.

Instead of black and white dots, shaded disks were used that elicit the perception of a concave dent or a convex bump depending on the direction of their shading (for a stimulus example, see Fig. 2.10). When the shading is light at the top and dark at the bottom, a convex bump is perceived and when the shading is inversed, a concave hollow is seen. The perception of depth is thought to occur due to the implicit assumption of a common light source that originates from overhead (Ramachandran, 1988). These "shape-from-shading" stimuli have been shown

to elicit robust texture segregation as well as "pop-out" in visual search (Ramachandran, 1988; Kleffner & Ramachandran, 1992; Kawabe & Miura, 2004) and it has been suggested that the 3D extraction occurs fairly early in visual processing and prior to perceptual grouping. Braun (1993) showed that shape-from-shading pop-out occurs even if attention is drawn to an additional task and suggested that it is thus attention-independent. Thus, perceptual grouping with such stimuli may occur relatively effortlessly (Kleffner & Ramachandran, 1992). Therefore, it was predicted that if the results of Exp. 1.1 can be replicated with shape-from-shading stimuli, then the effect of cueing might be less likely to have tapped perceptual grouping and more likely to have tapped numerosity extraction. However, if the effect of cueing disappears, than the experimental manipulation in Exp. 1.1. might have interfered with the grouping of stimuli into subsets rather than with enumeration.

2.3.2   Methods

Subjects

5 subjects (mean age: 29.6, range: 25-32, 2 females) were tested under the same conditions as in Exp. 1.1.

Stimuli

Shaded disks were generated by displaying a circular area from a smooth vertical grey scale shading (100 grey values) from white to black. Top-lit disks (convex bumps) were created with white on the top and black on the bottom, the inverse shading was used for bottom-lit disks (concave hollows). The greyscale gradient was linearised with the monitor's gamma value in order to correct for the non-linear output luminance of the screen. Disks had the same size as the dots in Exp. 1.1 (1°). Disks were drawn on a luminance-corrected exact mid-grey background, thus the overall luminance of each disk was the same as the background. In the stimulus, assignment of disk position on the screen, minimum distance between disks etc. was the same as in Exp. 1.1. The cue consisted of one centrally displayed convex or concave disk on a grey background. The mask consisted of concave and convex disks (50% each) drawn on
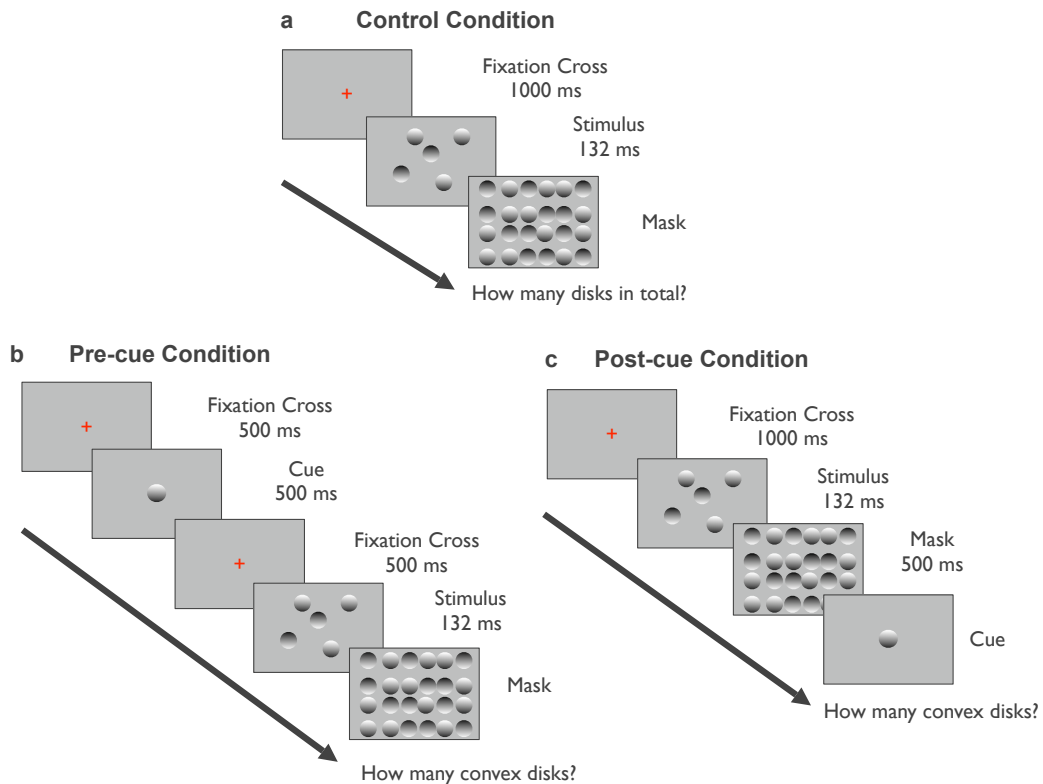
Figure 2.10: Experimental Procedure of Exp. 1.2.

a) Control condition b) Pre-cue condition c) Post-cue condition. Stimuli are drawn in a different scale than in the real stimulus.

a grid covering the whole screen. Each disk position was drawn with a slight random offset from the original grid position to create a randomly jittered pattern. The distribution of convex and concave disks whithin the grid was determined randomly in each trial anew. All other parameters were identical to Exp. 1.1.

Task, Experimental Procedure and Experimental Design

The stimulus was presented for 132 ms throughout. All other aspects of the task, the experimental procedure and the experimental design were identical to Exp. 1.1. See Fig. 2.10 for stimulus examples and the experimental procedure.

Data Analysis

When directly comparing results from dots and disks, data from Exp. 1.1. and Exp. 1.2. were merged in a repeated measures ANOVA with between-subject factor stimulus type.

2.3.3   Results

Overall, the results showed that in the cueing conditions, performance was markedly impaired compared to Exp. 1.1. This indicated that subjects were unable to perform the task, probably due to the complexity of the disk stimuli or due to insufficient stimulus presentation time, as discussed in more detail in Section 2.3.4. Given the limited conclusions that can be made from these findings, only significant results are reported.

Accuracy

Accuracy results are depicted in Fig. 2.11 with statistics in Table 2.9. With respect to the control condition, accuracy followed the same sigmoidal pattern as in Exp. 1.1. As would be expected, subjects performed very well in the low numerosities and became more error-prone in the high numerosities. In the cued conditions, however, performance differed markedly from the dot experiment. Although subjects performed on average above the defined chance level in the subitizing range (lowest t value: t = 2.34, highest p value $p = .079$, all other $p < .05$), accuracy was extremely low. In the counting range, performance did not depart significantly from chance (t ≤ .823, $p >.05$) or was at zero (pre-cue condition at numerosity 7 and both cued conditions at numerosity 9). In addition to the massive accuracy drop in the subitizing range, the difference between pre- and post-cue condition disappeared entirely and the overall effect of experimental condition was only due to the difference between the control and the cued conditions. The effect of target number disappeared in the subitizing range, indicating that accuracy was not better at numerosity 1 than at numerosity 4. The interaction between condition and target number remained, most likely due to the differential curves of control and cued conditions. Subjects reported not being able to distinguish the convex and concave disks from each other.
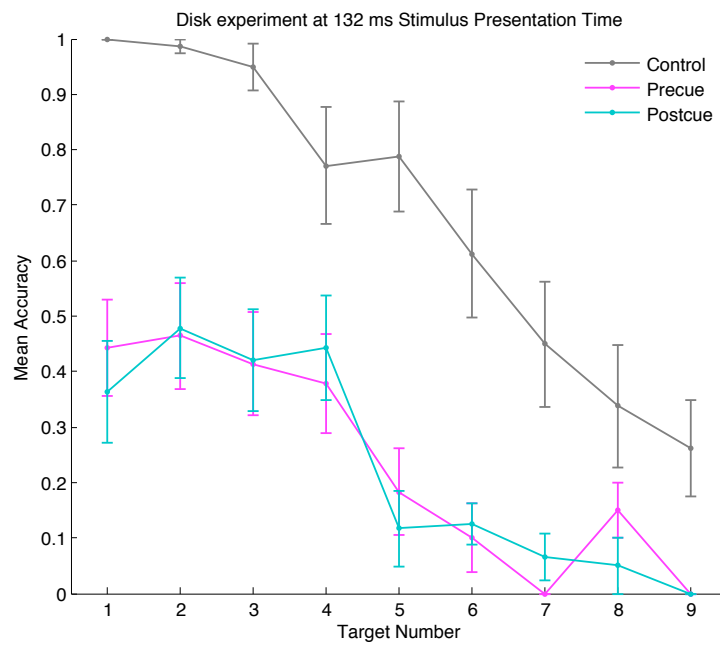
Figure 2.11: Exp. 1.2 Accuracy.

Accuracy data (proportion correct) as a function of target number for the disk experiment at 132 ms stimulus presentation time. $n = 5$. Note that the amount of trials differs between the control and the cued conditions. Error bars reflect 1 SEM.

When compared to the dot experiment (Fig. 2.12), performance in the cued conditions in the disk experiment was significantly worse than in the dot experiment in the subitizing range (between-subject factor stimulus type: $F(1,37) = 4.37$, $p = .044$). When taking the whole numerosity range or only the counting range into account this difference became non-significant ($F(1,37) \leq 2.23$, $p > .050$). Given that performance in the counting range was mostly at chance level in both the dot and the disk experiment, this is an unsurprising result. Pairwise post-hoc comparisons between experiments for each condition showed that in the subitizing range, performance with dots and disk differed significantly in the pre-cue condition ($F(1,37) = 5.45$, $p = .025$), but only marginally in the post-cue condition ($F(1,37) = 3.09$, $p = .087$). With respect to the control condition, performance was better in the disk than in the dot experiment ($F(1,37) = 9.70$, $p = .004$), mainly due to increased accuracy in the counting range ($F(1,37) = 14.6$, $p < .001$) rather than the subitizing range ($F(1,37) = 2.87$, $p > .050$). Note that stimulus presentation times were on average higher in the disk than in the dot experiment (except for those subjects tested at 150 ms).

Reaction Times

Given the chance level performance in the higher numerosities, reaction times for correct trials were only analysed for numerosities 1 - 5. The data mirrored the effects seen in accuracy data: cued conditions were responded to much slower than the control condition and the difference between pre- and post-cue condition disappeared (ANOVA results: main effect of condition: $F(2,8) = 15.66$ $p = .002$, post-hoc comparisons: control versus pre-cue: $p = .010$, control versus post-cue $p = .014$, pre- versus post-cue: $p > .050$). There was a main effect of target number ($F(4,16) = 9.00$, $p = .001$) but no interaction ($F(8,32) = 1.09$, $p > .050$). Results replicated when considering only the subitizing range from 1-4 except for the difference between control and pre-cue condition ($p = .060$). Mean reaction time data is plotted as a function of target number in Fig. 2.13.

When considering all experimental conditions (Fig. 2.14), RT in the disk experiment did not differ from the dot experiment (between-subject factor stimulus type: $F(1,37) = 1.58$, $p > .050$) but there was an interaction of condition with stimulus type ($F(2,74) = 20.66$, $p <$
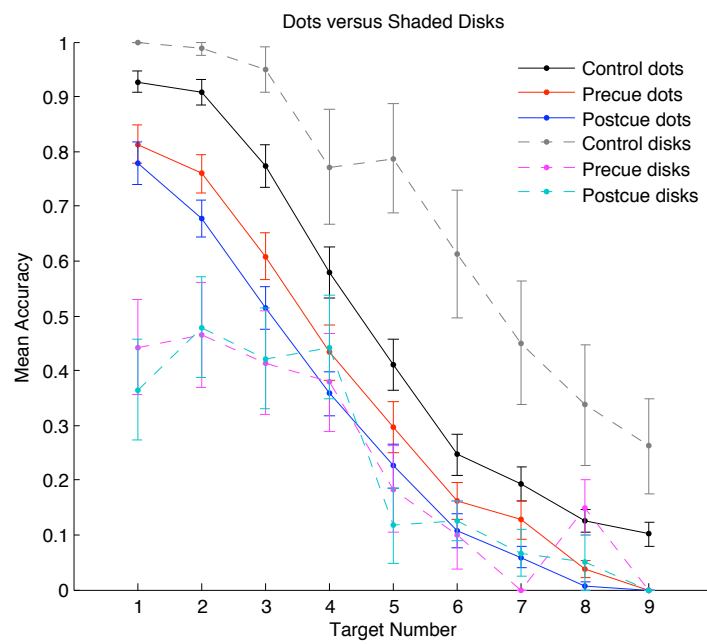
Figure 2.12: Exp. 1.2 Accuracy - Dots versus Disks.

Accuracy data for both the disk and the dot experiment (pooled across all presentation times). Dots: $n = 34$; Disks: $n = 5$. Error bars reflect 1 SEM.

Exp. 1.2 Accuracy - Shaded Disks at 132 ms Presentation Time

		Numerosity Range		
		1 - 9	1 - 4	5 - 9
Within-Subject Effects				
Condition	p	<.001	<.001	<.001
	F(df)	54.14 (2,8)	109.8 (2,8)	24.5 (2,8)
Target Number	p	<.001	n. s.	.001
	F(df)	32.87 (8,32)	2.94 (3,12)	7.54 (4,16)
Condition x Target Number	p	.012	.043	.053
	F(df)	2.25 (16,64)	2.62 (6,34)	2.22 (8,32)
Post-hoc Comparisons				
Condition precue vs postcue - planned	p	n. s.	n. s.	n. s.
control vs precue	p	.008	.001	.030
control vs postcue	p	.003	<.001	.017

Table 2.9: Exp. 1.2 Accuracy - Disk Experiment

Repeated Measurement ANOVA results for accuracy data of Exp. 1.2. using shaded disks at 132 ms stimulus presentation time. $n = 5$. Non-significant effects indicate $p > .05$. p values in post-hoc comparisons are Bonferroni-corrected except where planned.
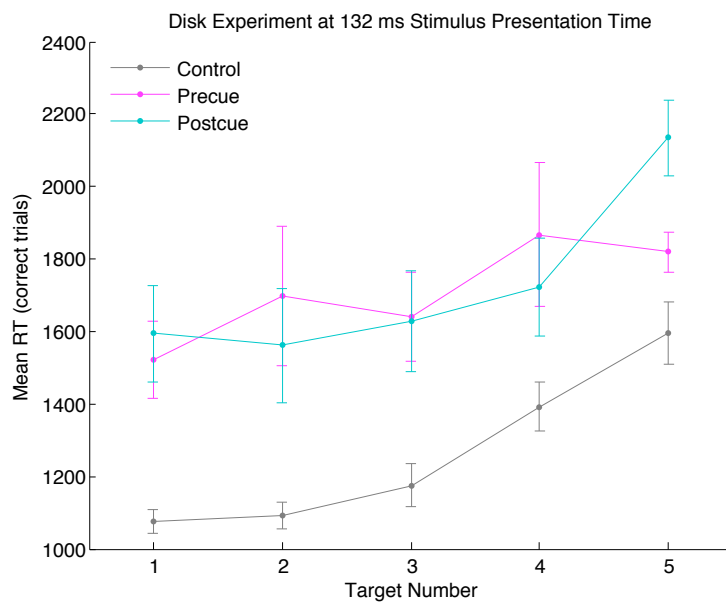
Figure 2.13: Exp. 1.2 Reaction Times.

Reaction time data (correct trials only, in ms) as a function of target number for the disk experiment at 132 ms stimulus presentation time. $n = 5$. Error bars reflect 1 SEM.
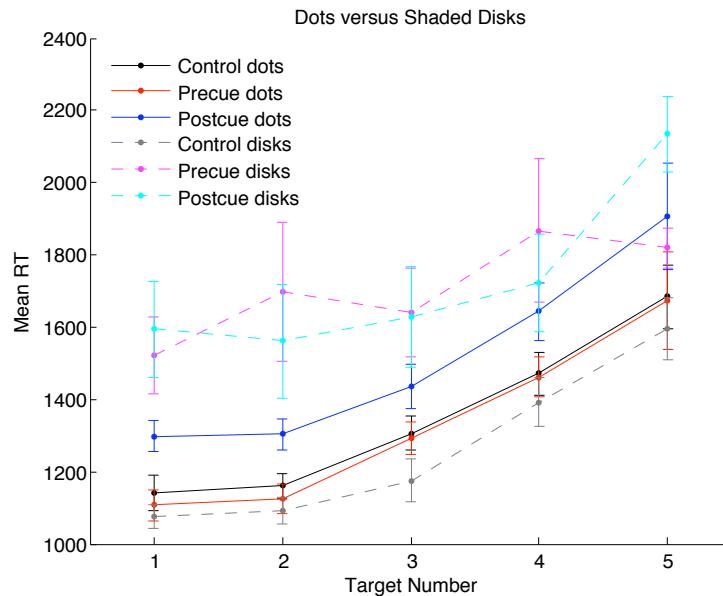
Figure 2.14: Exp. 1.2 Reaction Times - Dots versus Disks.

Reaction time data for both the disk and the dot experiment (pooled across all presentation times). Dots: $n = 34$; Disks: $n = 5$. Error bars reflect 1 SEM.

.001). When comparing conditions separately across experiments, RT in the pre-cue condition was significantly higher in the disk than in the dot experiment F$(1,37) = 7.49$, $p = .009$). Effects replicated in the 1-4 range.

Mean Responses, Response Variance and Weber Fraction

Mean responses and response variances are plotted in Fig. 2.15 and Fig. 2.16. Mean responses mirrored the accuracy data: performance in the control condition was very good, whereas it was extremely bad in the cued condition. Mean response curves for the cued conditions were much flatter than the control condition, indicating that subjects spread their responses similarly across numerosities. Performance in the control condition did not differ from perfect performance

except of a slight underestimation at numerosity 9 (t = -3.67, p = .021; all other numerosities t ≤ 2.75, p > .050). In the cued conditions however, mean responses departed from correct response in almost all numerosities except for numerosities 3 and 4 (t ≤ -2.24, p > .050, all other numerosities t ≥ 3.05, p ≤ .038). As accuracy at target number 3 and 4 was very low, this only indicates that over- and underestimation were balanced. At numerosities lower than 3, overestimation prevailed, whereas from numerosity 4 onwards, underestimation prevailed.

There was a main effect of condition due to the difference between control and cued conditions. However, the cued conditions did not differ from each other, mirroring the accuracy data. Unsurprisingly, there was a main effect of target number (F(8,32) = 49.6, p < .001) that interacted with condition (F(16,64) = 16.92, p < .001).

Response variance data was in line with the observations of accuracy and mean responses. Response variances increased with target number only in the control condition (F(8,32) = 8.56, p <.001). The lack of an effect of target number indicates that response behaviour varied similarly across all numerosities. There was a main effect of condition, again mainly driven by differences between the control and the cued conditions but without difference between cued conditions (Main effect: F(2,8) = 35.6, p < .001; control versus pre-cue: p = .007; control versus post-cue: p = .005, pre- versus post-cue: p > .050).

When comparing mean responses between the disk and the dot experiment, the between subject effect was not significant (F(1,37) = .67, p >.050), but there was a significant interaction between stimulus type and experimental conditions (F(2,74)=8.68, p <.001). When comparing conditions separately across experiments, mean responses differed only in the control condition (F(1,37) = 4.98, p = .032). Response variances did not differ between experiments, however, there was a significant stimulus type - condition interaction (F(2,74) = 6.64, p = .002).

Weber fraction showed a similar pattern as in Exp. 1.1. and are plotted in Fig. 2.17. There was a main effect of condition (F(2,8) = 137.24, p < .001) which was due to a significant increase in Weber fraction in both cued conditions compared to the control condition (p < .001 in both post-hoc comparisons). There was a main effect of target number (F(8,32) = 88.15, p < .001) which replicated when each condition was analysed separately (all p < .001). Target number and condition interacted (F(16,64) = 59.85, p < .001). All effects replicated when

analysing the subitizing and counting ranges separately.

Overall, Weber fraction did not differ between experiments. However, when comparing conditions separately with each other, Weber fraction in the pre-cue condition of the disk experiment was significantly higher than in the pre-cue condition of the dot experiment (F(1,37) = 4.74, p = .036). The difference between post-cue conditions across experiments was only marginally significant (F(1,37) = 3.77, p = .060).
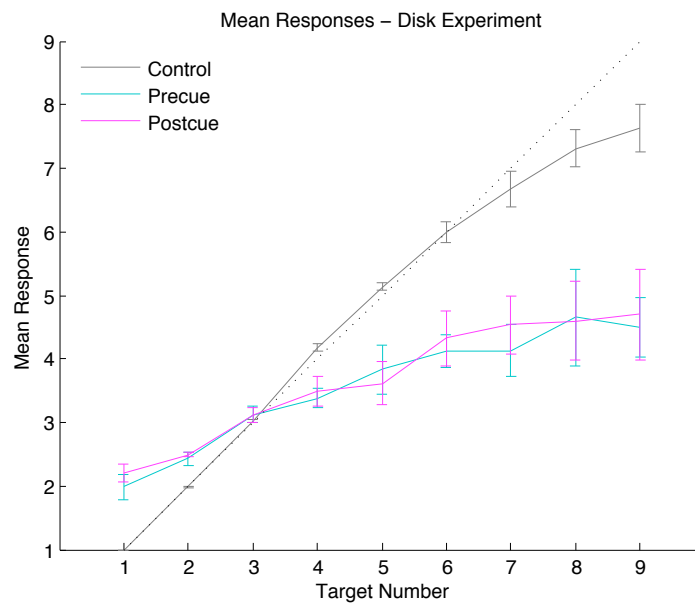
Figure 2.15: Exp. 1.2 Mean Response.

Mean response as a function of target number for the disk experiment at 132 ms stimulus presentation times. $n = 5$. The dotted line indicates perfect performance. Error bars reflect 1 SEM.
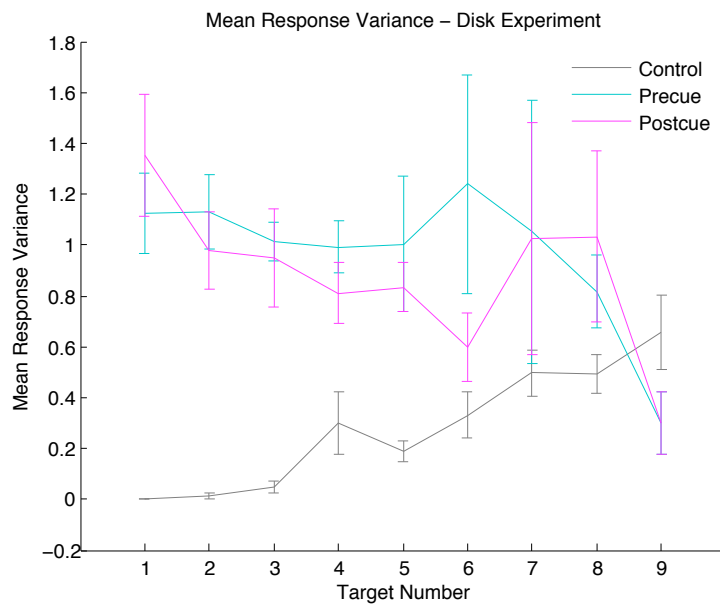
Figure 2.16: Exp. 1.2 Mean Response Variance.

Mean response variance as a function of target number for the disk experiment at 132 ms stimulus presentation time. $n = 5$. Error bars reflect 1 SEM.
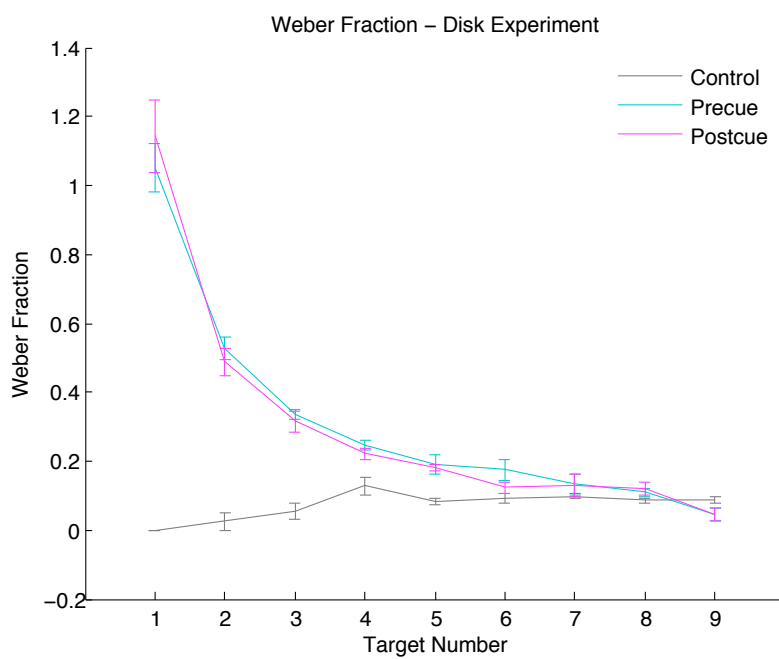
Figure 2.17: Exp. 1.1 Weber Fraction.

Weber fraction (response standard deviation / target number) as a function of target number for the disk experiment at 132 ms stimulus presentation time. $n = 5$. Error bars reflect 1 SEM.

Effects of Stimulus Heterogeneity

Visual search studies using shape-from-shading stimuli have demonstrated search asymmetries between concave and convex disks. Depending on the stimulus settings and the experience of subjects to view such stimuli, concave disks may produce visual "pop-out" among convex disks but not vice versa (Kleffner & Ramachandran, 1992) or convex disks may pop out from concave disks but not vice versa (Kawabe & Miura, 2004). Given these finding, there might be similar asymmetries between convex or concave disks in enumeration. Therefore, the data from this experiment were separated according to shading type and analysed for differences. No significant difference in enumeration performance between convex and concave disks was found, neither in accuracy nor RT data, and neither in the pre-cue nor in the post-cue condition. Data in the control condition were analysed according to uniform (only convex or only concave disks) and mixed (both convex and concave disks) displays, but as in the dot experiment, no difference in neither accuracy nor RT data was found (all p values above .050).

2.3.4   Discussion

Changing the stimuli from black and white dots to shaded disks had a rather surprising effect: when asked to subitize only a subset of the stimuli, subjects gave highly inaccurate, slow and rather random responses. As participants reported subjectively not being able to distinguish the shaded disks from each other, it appears that a floor effect occurred. Even being able to direct attention to one type of stimulus and ignore the other (pre-cueing) did not bear any behavioural advantage any more to the situation when both types of stimuli could be relevant (post-cueing). Furthermore, the lack of an effect of target number in the subitizing range indicates that 1 item was not any more easier to enumerate than 4 items. This was true despite relatively high stimulus presentation times compared to the dot experiment.

Taken together, these results strongly suggest that subjects were unable to group convex and concave disks into different subsets and enumerate them within 132 ms. When grouping of the items into different subsets was not required to perform the task, as it was the case in the control condition, subitizing performance was excellent and, due to the relatively high

stimulus presentation times, better than in the dot experiment. The almost perfect subitizing performance in the control condition is supported by both accuracy, reaction time, mean response and response variance data. This suggests that shaded disks could easily be segregated from the background and enumerated. The limiting factor in the cued conditions must have therefore been either the discrimination of differently shaded disks from each other or their grouping into subsets on the basis of perceived depth. This interpretation is supported by high Weber fractions and thus low discriminability in the cued conditions compared to the control condition and compared to the dot experiment. Experiment 1.3. was carried out in order to address the issue of discrimination experimentally.

Possible Criticisms

It is well possible that the emergence of perceived depth from 2D shading and/or the discrimination and grouping of shaded disk might require longer stimulus exposure than that employed in the current experiment. In fact, previous studies using shaded disks have employed unlimited or relatively long exposure times (1 second or longer) and hardly looked at shape-from-shading perception in short masked presentations (Ramachandran, 1988; Kleffner & Ramachandran, 1992). Kawabe and Miura (2004) were the first to study figure-ground segregation with shaded disks and short presentation times (50, 250, 500 and 1000 ms). They presented a windmill shape composed of convex and concave disks and asked subjects whether the vanes of the windmill were right- or left- tilted. When the disks were slightly spatially separated from each other, figure-ground segregation was at chance for stimulus presentation times below 500 ms, supporting the idea that concave and convex disks might not be discriminated or grouped at short presentation times. However, when disks were arranged immediately adjacent to each other, figure-ground segregation was above chance performance also at shorter presentation times. Here, when a vane of convex disks neighboured a vane of concave disk, a clear contour emerges and the task could be done on the basis of texture segmentation and contour without the need to discriminate or group the disks on the basis of their perceived shape.

Trick and Enns (1997) used shaded disks in an enumeration task. Here, subjects were able to enumerate convex amongst concave shapes and vice versa. However, the stimulus was

displayed until subjects responded (up to 7.7 seconds) and verbal counting was not prevented. High reaction times (ranging from 600 to 3000 ms) indicate that discrimination, grouping and enumeration of targets took up some processing time and might require prolonged stimulus exposure. Even with the long stimulus exposure that Trick and Enns employed, the reaction time slopes for enumeration were higher than would be expected for stimuli that produce rapid texture segmentation, thus indicating that some processes enumeration might be contingent on caused the slowing down of responses.

Thus, it is likely that with higher stimulus presentation times, subjects might have been able to perform the task in the current experiment. However, as mentioned earlier, presentation times above 150 or 200 ms allow eye movements and do not prevent serial counting anymore and are therefore inappropriate to study immediate subitizing.

Implications for Perceptual Grouping on the Basis of Shape-from-Shading

The results have a further implication: if the discrimination and grouping of disks on the basis of their perceived depth could not be achieved within the employed stimulus presentation time, then this suggests that perceptual grouping by perceived depth is not as effortless as previously assumed. This speaks against the conclusion by Kleffner and Ramachandran (1992) and Braun (1993) that the extraction of shape from shading occurs early in visual processing. That grouping by perceived depth is more difficult than grouping by luminance might not be such a surprising finding: the shaded disks are more complex stimuli than the dots, they are equiluminant and differ from each other only in the direction of their shading. Extraction of depth from shading might require time, thus the discrimination of convex from concave disks is a more subtle and probably more processing demanding discrimination than that of luminance (particularly as luminance differences were chosen to be as large as possible). In fact, with respect to their strength as a grouping cue, shape from shading stimuli have never been directly compared to stimuli differing in luminance or colour.

Therefore, the reason for the discrepancy between the current results and the conclusions made by Kleffner and Ramachandran (1992) might lie in the nature of comparison. Kleffner and Ramachandran (1992) compared vertically shaded disks with their ambiguous horizontally

shaded counterparts and with bi-partite disks (disks half filled with black and half with white without any shading). In their segregation and motion detection task, vertically shaded disks (those that elicit unambiguous shapes) had a consistent advantage over the ambiguous disks and the bi-partite disks that did not elicit perceived depth. The authors concluded that shape-from-shading extraction can drive segregation and motion discrimination and should therefore occur relatively early in visual processing. However, exposure duration was again relatively long and thus makes it difficult to judge the time course of a certain process. Moreover, this conclusion was made in comparison to equally complex and, in the case of horizontally shaded disks, perceptually ambiguous stimuli. It is well possible that simple luminance stimuli, for example, would have yielded much better performance in comparison to shaded disks.

Stimulus Induced Asymmetries

In the cued conditions, there was no effect of stimulus type (convex versus concave disks) in enumeration performance. That means convex disks amongst concave disks were not enumerated faster or more accurately than the opposite combination. This is in line with the findings of Trick and Enns (1997) and suggests that enumeration does not yield a similar asymmetry than visual search or texture segmentation. It has to be noted though that such asymmetries were often observed in stimulus arrays with stimulus grids of closely adjacent shaded disks (Kleffner & Ramachandran, 1992; Kawabe & Miura, 2004) and thus might reflect more of an effect of segmenting whole areas with similar textures from each other. In the current experiment, however, shaded disk were separated from each other by at least one dot diameter of grey background and did deliberately not fall into a regular grid pattern. In line with the results of the windmill task by Kawabe and Miura (2004) reported above, this might have removed any effect of texture segmentation.

The lack of an effect of stimulus heterogeneity (uniform versus mixed stimuli) in the control condition confirms the results from the dot experiment and suggests that subitizing performance is neither hindered nor promoted by task irrelevant feature differences.

The Problem of Defining Chance Performance

In this experiment, although subitizing accuracy was very low, it was on average above the here defined chance level. However, care must be taken with the definition of chance performance in this experiment and in enumeration experiments in general. Even if adjusted for trial frequencies, it is difficult to determine chance performance as subjects are unlikely to respond in an uniformly random manner across the whole range of response possibilities when they do not know the correct answer. It is more likely (and often observed in individual data) that subjects choose a preferred numerosity (e.g. 3) whenever they are unsure. Evidence for this tendency is more clearly apparent in accuracy and response variance data of Experiment 2 (Chapter 3). It might also be the case that subjects on average deviate towards the middle of the given numerosity range, possibly by excluding the extremes. Alternatively, the often observed so-called end-effect (i.e. reaction time drop /accuracy rise at the highest numerosity) indicates that subjects sometimes employ strategies and deliberately choose the extremes whenever they see very few or very many targets present. In the case where exact enumeration fails, subjects might still have an approximate idea whether the number of targets was rather on the low or on the high end of the scale of response possibilities and might be able to give a rough guess. As discussed earlier, the probabilistic distribution of possibilities to make an incorrect judgement varies with target number. That is, at target number 1 the brain would have to "make up" a signal in order to make a largely incorrect numerosity judgement (say 9) which would be relatively unlikely to occur. At higher numerosities, however, the whole range of lower numerosities can be chosen as incorrect response just by missing out some of the signal. In addition, the probability of choosing one of those lower numerosities varies, e.g. at target number 9, it is more likely that subjects judge 7 or 8 rather than 1 or 2.

Taken together, all those possibilities can bias response behaviour to a non-uniform distribution and thus make it very difficult to determine how subjects respond when they have to rely on guessing.

Conclusion

With respect to the original hypothesis, instead of facilitating discrimination and/ or perceptual grouping, shape-from-shading stimuli seem to hinder such processes at the employed stimulus exposure time resulting in a failure of enumeration altogether. This suggests that target-distractor discrimination or perceptual grouping might be crucial processes that enumeration is contingent on. This is further evidence against a pre-attentive subitizing mechanism.[2]

2.4   Experiment 1.3 Discrimination of Shaded Disks and Dots

2.4.1   Rationale

Based on the results of Experiment 1.2., this experiment aimed to address the question whether subjects are able to detect and discriminate convex from concave disks under the employed short stimulus exposure. The experiment was also run with black and white dots in order to compare discrimination difficulty between the two stimulus types.

Subjects were presented with displays containing only one type of stimulus or a mixture of both and were asked to determine which of three alternatives the display was composed of: either only convex disks or only concave disks or both. The total number of items in the display varied as well as the proportion of both stimulus types in the mixed condition. It was hypothesised that this task required to detect the items, determining their identity (and presumably extracting 3D shape from 2D shading) and discriminating them from each other. As the mixed condition sometimes contained only one item from one type and many (up to 9) items from the other type, the task required the detection and identification of all items in the display.

---

[2]In addition to the current experiment, a disparity version of this experiment was carried out. Gabor patches were displayed both in the front and in the back plane using shutter glasses. The cue consisted of a ring that moved from the middle plane either towards the front plane or towards the back plane and back again to the middle plane. Thus, cueing occurred on the level of true depth rather than on the level of luminance or perceived depth from shade-from-shading. It turned out that it was impossible for subjects to distinguish front items from back items and enumerate them separately, even with an increased stimulus presentation of 250 ms (which is already beyond the rationale of this paradigm). 3 subjects were tested and performance was entirely at chance.

It was predicted that if detection and discrimination of items under the employed presentation times were the limiting factors in the enumeration experiments 1.1. and 1.2., then performance should be relatively low, i.e. at or near chance level. However, if discrimination performance is high, a failure in item detection and discrimination in those previous experiments could be ruled out. Furthermore, performance with shaded disks should be much worse than with black and white dots, if a failure in detection and discrimination causes the difference between Experiment 1.1 and Experiment 1.2. However, if performance between stimulus types does not differ, the differential results between both enumeration experiments cannot be ascribed to a difference in stimulus-evoked discrimination difficulty.

2.4.2   Methods

Subjects

6 subjects were tested with shaded disks (mean age: 27.6, range: 24-32, 4 females) and 5 subjects (mean age: 27.0, range: 24-29, 3 females) with black and white dots. 3 subjects participated in both experiments. Testing conditions were the same as in Exp. 1.1. and 1.2.

Stimuli

Stimuli characteristics were the same as in Exp. 1.1. and Exp. 1.2., respectively.

Task and Experimental Procedure

Two versions of the experiment were run, one with dots and one with shaded disks. Three experimental conditions were employed: two displaying only one alternative of the stimulus (only black and only white (dot version), only convex and only concave disks (disk version) and one displaying a mix of both (black and white dots/ convex and concave disks). The experimental procedure is shown in Fig. 2.18. Subjects viewed first a fixation cross (1000 ms), then the stimulus (dots: 50 ms, disks: 132 ms) followed by a mask (the screen filled with dots or disks). The mask stayed on the screen until subjects responded. Subjects were asked to
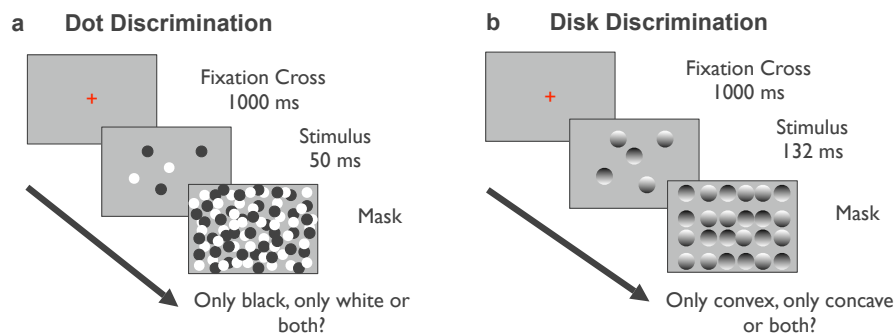
Figure 2.18: Experimental Procedure of Exp. 1.3.

a) Version with black and white dots b) Version with shaded disks. Stimuli are drawn in a different scale than in the real stimulus.

decide whether they had seen only black dots / concave disks, only white dots / convex disks or both black and white/convex and concave. Accuracy was emphasised. Subjects performed one practice block before the two experimental blocks. The testing session lasted for about 15 min.

Experimental Design

Subjects performed two blocks of 60 trials each, each block comprising 20 trials for each condition (40 trials for each condition in total). Total number of items in the stimulus ranged from 2 to 10 and was counterbalanced. In the mixed condition, all possible combinations to divide the stimuli into two subsets were used and counterbalanced as far as possible. I.e. in some trials, only one item was of one kind and all others were of the other kind, whereas sometimes both

kinds were more equally distributed.

Data Analysis

Accuracy and reaction time data were analysed with a repeated measure ANOVA with within subject factor condition (3 levels) for each stimulus type separately. Data were combined to determine differences between types. α levels for post-hoc comparisons were corrected after Bonferroni. As reaction times were measured from stimulus onset on, reaction times with shaded disks were adjusted (- 82 ms) when compared with dots.

2.4.3   Results

Accuracy

Mean accuracy for all three conditions is plotted in Fig. 2.19 for shaded disks and for black and white dots. For the shaded disks, accuracy overall was high (mean = 0.814, SEM = .058). Accuracy differed between the three conditions (main effect: $F(2,10) = 8.10$, $p = .008$) which was due to lower accuracy in the condition with only concave disks compared to only convex disks ($p = .022$). All other post-hoc comparisons were non-significant ($p > .050$). For the dots, accuracy overall was at 0.79 (SEM = 0.059), with a significant main effect of condition ($F(2,8) = 5.39$, $p = .033$). Post-hoc comparisons revealed a marginally significant difference between the only black and the mixed condition ($p = .058$) but no difference between the other conditions ($p > .050$). When analysing dots and disks together, no between-subject effect was found ($F(1,9) = .21$, $p > .050$), but an interaction of condition with stimulus type ($F(2,18) = 5.06$, $p = .018$).

Reaction Times

Mean reaction time for shaded disk discrimination was 1210 ms (SEM = 108.6) without any significant differences between conditions (only concave disks: mean RT = 1305.6 ms (SEM = 116.7); only convex disks: mean RT = 1168.3 ms (SEM = 113.2); both: 1156.0 ms (SEM = 95.9); $F(2,10) = 1.54$, $p > .050$). Black and white dots yielded slightly shorter reaction times
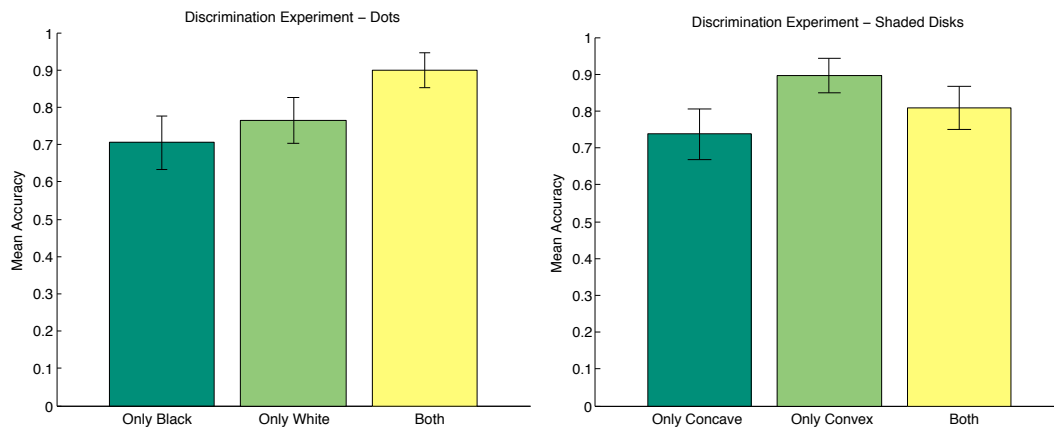
Figure 2.19: Exp. 1.3 Accuracy.

Accuracy data of the discrimination experiment with black and white dots (left) and shaded disks (right) for the three experimental conditions. Dots were displayed for 50 ms ($n = 5$), shaded disks for 132 ms ($n = 6$). Error bars reflect 1 SEM.

(overall 983.1 ms (SEM = 65.7), and again no effect of conditions (only black dots: mean RT = 996.5 ms (SEM = 68.5); only white dots: mean RT = 1000.8 ms (SEM = 69.9); both: 951.9 ms (SEM = 58.6); $F(2,8) = .54$, $p > .05$)). Reaction times to dots did not differ from reaction times to shaded disks (between-subject effect: $F(1,9) = .87$, $p > .05$). There was no interaction between condition and stimulus type ($F(2,18) = .82$, $p > .05$).

2.4.4   Discussion

The results of this experiment clearly showed that subjects were very well able to detect, identify and discriminate the different types of stimuli from each other at the stimulus presentation times employed for the enumeration experiments. Accuracy was high and did not differ between shaded disks and black and white dots. Also reaction times did not indicate differential processing time for both stimulus types. This implies that the failure to enumerate convex disks amongst concave disks or vice versa was not due to the inability to identify and discriminate the shaded disks from each other. As discrimination performance was comparable with black and white dots, the enumeration advantage of dots must lie within a process different from (and possibly later than) detection and discrimination. Apart from enumeration, perceptual grouping might be the factor that could be affected by the cueing manipulations of the current paradigm, given the reasoning in earlier sections. Experiment 1.4 was designed to address this issue.

The results are in accordance with Atkinson, Campbell and Francis (1976) who showed that, at low spatial frequencies, items can be well resolved and discriminated but much less easily enumerated (only up to 2 items could be enumerated without errors). They suggest that a process of perceptual segregation is necessary for enumeration, very much in line with the here proposed interpretation.

Stimulus Induced Asymmetries

In this task, some of the asymmetries between stimuli observed in earlier studies became apparent. Determining whether the whole display was composed of convex disks was achieved with more accuracy than when only concave disks were present. This might be in line with the finding

by Kawabe and Miura (2004) showing that figures composed of convex disks are perceived more readily as emerging figure than concave disks. However, tasks and stimuli characteristics were different in the present study and the one by Kawabe and Miura, so care must be taken with a direct comparison. Nevertheless, performance asymmetries between convex and concave disks is a common finding (Kleffner & Ramachandran, 1992). With black and white dots, performance asymmetries were slightly different. Determining mixed displays could be done with higher accuracy than determining uni-coloured displays (particular only black), but without conclusive differences in post-hoc comparisons between conditions.

Possible Criticisms

It remains unclear whether the disk discrimination task was performed truly on the basis of perceived 3D depth or on the basis of the direction of shading. To decide whether the whole display was composed of one type of disk or both, subjects could have monitored whether the shading direction was uniform or not. In order to identify convex or concave disks the direction of shading had to be assigned to one stimulus type. Therefore, even if the 3D shape perception could not be achieved at the employed exposure time, this discrimination task could have been carried out on the basis of shading direction. In order to control for this confound, bi-partite disks (half white, half black disks without shading gradient) could be used similar to those in previous studies (Kleffner & Ramachandran, 1992). As the behavioural effects with shaded disks normally disappear with bi-partite disks, it has commonly been assumed that it is perceived depth that underlies such effects (Kleffner & Ramachandran, 1992). However, even if this current discrimination task could have been carried out on the basis of shading, enumeration might have been carried out on the same basis, too. Regardless of whether the discrimination of convex from concave disks is done on the basis of shading direction or perceived depth, selecting a subset and enumerating it appears to require prolonged exposure and possibly attentional resources.

2.5   Experiment 1.4 Effects of Attentional Set on Visual Search

2.5.1   Rationale

Based on the results of experiments 1.2 and 1.3, it might have been the perceptual grouping of targets and the selection of a target set rather than enumeration that was affected by the cueing manipulations. In order to address this possibility, a visual search experiment was designed that resembled the enumeration task as closely as possible with respect to requirements of grouping and target set selection but without involving a judgement of numerosity or quantity. Furthermore, a search task traditionally defined as pre-attentive was chosen in order to match the supposed attentional requirements of subitizing. Search for a horizontal line amongst vertical distractors or vice versa has been shown previously to be a classical feature search or "pop-out" and thus supposedly a pre-attentive task (Sagi & Julesz, 1985; Treisman & Gormican, 1988). Thus, an array of black and white horizontal and vertical lines was used and subjects were asked to detect the odd-ball differing in orientation from all other items. It was hypothesised that detection of an orientation odd-ball requires perceptual grouping as the common orientation of all items needs to be determined first in order to identify a deviating item.

The control condition was a classical visual search paradigm where subjects were asked to detect the orientation odd-ball while ignoring the colour of the stimuli. The colour distribution as well as both the number of items in total and those in the subsets were chosen as similarly as possible to the enumeration experiment. As in a typical visual search experiment, half of the trials contained a target whereas the other half did not.

In the cued conditions, however, every trial contained a target and thus the task resembled more a detection task. Here, subjects were asked to select the cued subset and decide whether the target was contained in the cued subset or not. As in the enumeration experiment, pre-cueing posed an attentional advantage towards the subset to select, whereas post-cueing did not, i.e. both subsets needed to be attended to. The numerosity of the subsets to select were matched as closely as possible to the enumeration experiment in order to make both experiments as similar in perceptual requirements as possible. It has to be noted that there was an alternative way of doing the task: subjects could first search for the orientation odd-ball and

determine then whether the colour matched the cue (without necessarily having to select the whole subset). In theory, however, this was true for the enumeration experiment, too. Subjects could enumerate first both subsets and then report the one that matched the cue colour.

It has to be pointed out that although subjects had to look for an orientation odd-ball of a specific colour in the current task, this was not a conjunction search. In such search, the target is defined by a conjunction of features (e.g. a horizontal white bar) and has to be distinguished from distractors that share one of each feature (e.g. horizontal black bars and white vertical bars) (e.g Treisman & Gelade, 1980). In this task, however, the target just shared one feature, namely colour, with some of the distractors (those of the same subset), but not orientation.

It was hypothesised that if cueing affects either the perceptual grouping of targets or the selection of the target set, a similar pattern than in the enumeration experiment would be expected: subjects should be less accurate and slower in the post-cue condition than in the pre-cue condition. However, if solely enumeration was affected by cueing, then no difference between the cued conditions should be expected in the current search task.

The "pre-attentive" nature of the search task was assessed by analysing performance as a function of set size. A flat curve, i.e. no increase of reaction times or error rates with the number of items present should indicate parallel feature search as traditionally defined (Treisman & Gelade, 1980). A sloped curve, however, would indicate serial search. Classical visual search paradigms most often employed unlimited stimulus exposure and used mainly reaction times as dependent measure. As stimulus exposure was limited in the current paradigm, both reaction times and accuracy were analysed with respect to set size.

2.5.2 Methods

Subjects

9 subjects participated (mean age: 25.6, range: 21-38, 5 females). Testing conditions were the same as in the previous experiments.
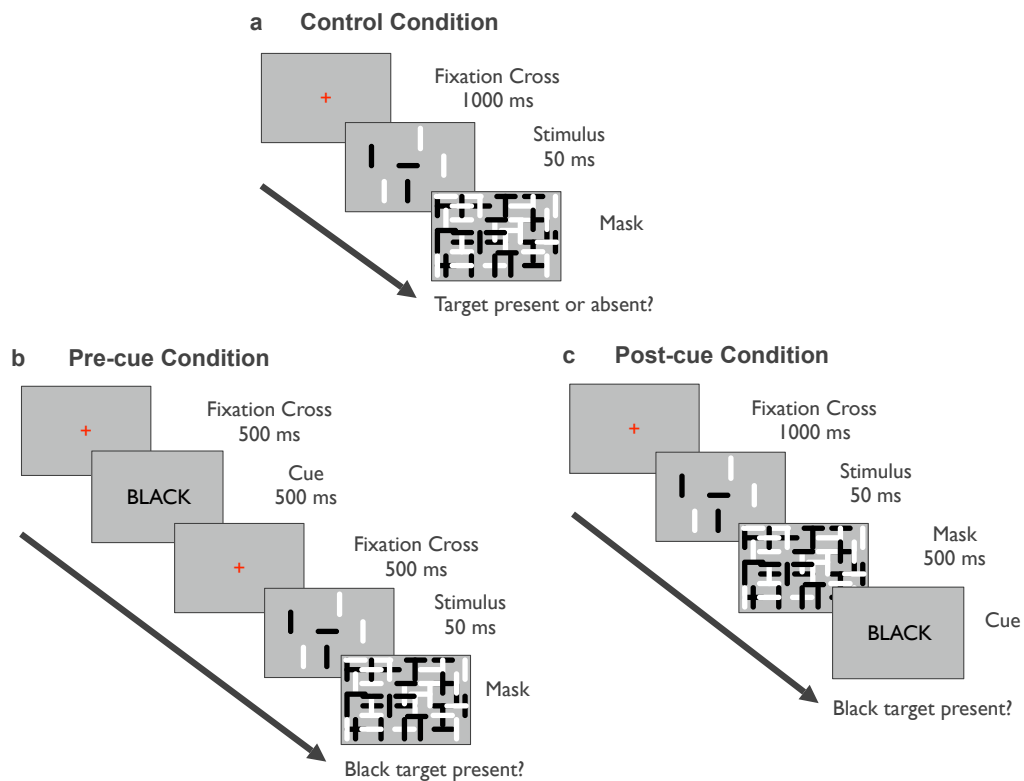
Figure 2.20: Experimental Procedure of Exp. 1.4.

a) Control condition b) Pre-cue condition c) Post-cue condition. Stimuli are drawn in a different scale than in the real stimulus.

Stimuli

The stimulus consisted of black and white bars ($1.3°$), either vertically or horizontally oriented. Bars were displayed on a restricted central field identical to the previous experiments. Randomised position on the display, minimum distance between stimuli, background and luminance adjustment were the same as in the other experiments. The cue was the same as in Exp. 1.1., i.e. the words "BLACK" or "WHITE" displayed in the respective colours. The mask consisted of 1000 black and white bars, randomly intermixed, partly overlapping and covering the whole screen. See Fig. 2.20 for a stimulus example.

Task and Experimental Procedure

Subjects were asked to detect the odd-out item in the display, i.e. the bar that differed in orientation from all other bars. In the control condition subjects detected the presence or absence of the odd-ball while ignoring the colour of the stimulus. In the cued-condition, subjects were asked to detect an orientation odd-ball in the cued subset only. Here, there was always an orientation odd-ball present in the display, subjects were required to decide whether the target belonged to the cued subset or not. The stimulus was always displayed for 50 ms, all other aspects of the experimental procedure were identical to Exp. 1.1 and 1.2 (see Fig. 2.20).

Subjects responded as accurately as possible on two adjacent keys on the keyboard, indicating target presence or absence (in the control condition for the whole display, and in the cued conditions for the cued subset). Subjects received practice trials before each condition and were encouraged to take a break after half of the trials in each block. The testing session lasted for about 45 minutes.

Experimental Design

Subjects performed 112 trials in each condition (336 in total). In the control condition, half of the trials were uni-coloured (half black, half white) and half of the trials were mixed coloured. In each combination, half of the trials contained a target and half did not. Orientation of bars was counterbalanced (i.e. in each combination, half of the trials contained horizontal bars with vertical odd-ball (if present) and the other half contained the opposite orientation). Total number of items ranged from 4-10 and was again counterbalanced for orientation, target presence and colour.

In the cued conditions, the display was always mixed coloured and an orientation odd-ball was always present. Half of the trials required to select the black and half of them the white subset, and accordingly, half of the trials contained the target in the cued subset and half of them contained the target in the distractor subset. The total number of items ranged from 4-10 and the number of items in the colour subsets ranged from 2-8. The total number of items could not be entirely counterbalanced, but the frequency to select a subset with a specific number of

items was matched to the trial frequencies of the enumeration experiment. For example, the frequency with which the black subset containing 3 items needed to be selected was proportional to the trial frequency with which 3 black items needed to be enumerated in Exp. 1.1. As in the enumeration experiment, larger subsets of 5 items and above were required to be selected less frequently than smaller subsets. I.e. in the 56 trials where black was the cued subset, 42 trials contained a black subset of 2, 3 or 4 items (14 trials each), 6 trials contained 5 items, 4 trials contained 6 items and 4 trials contained 7 or 8 items (2 trials each). Naturally, the same was true for the case when the white subset was cued. Consequently, the proportion between the number of items in target subset and distractor subset was proportional to the enumeration experiment. As in the control condition, orientation of bars was counterbalanced for each combination as far as possible.

Data Analysis

Accuracy and reaction times were analysed in dependence of several variables. In the control condition, these were target presence, total number of items in the display and colour heterogeneity (uni-coloured or mixed-coloured). The cued conditions were analysed with respect to the total number of items, the number of the target subset and the number of the distractor subset for each condition separately. Overall performance in the cued conditions was compared with each other and with the control condition. Repeated-measure ANOVAs were used in all cases. Reaction times are reported for correct trials only.

2.5.3   Results

From the 9 tested subjects, 2 subjects performed below chance level (0.5) and 2 performed close to chance level (0.52 and 0.53, respectively) in the cued conditions. All 4 were excluded from the analysis so that the data of 5 subjects remained.

Overall, mean accuracy was highest in the control condition and lowest in the post-cue condition (see Fig.2.21). There was a main effect of condition ($F(2,8) = 12.9$, $p = .003$) where the control condition did not differ from the pre-cue condition ($p >.05$) but almost differed
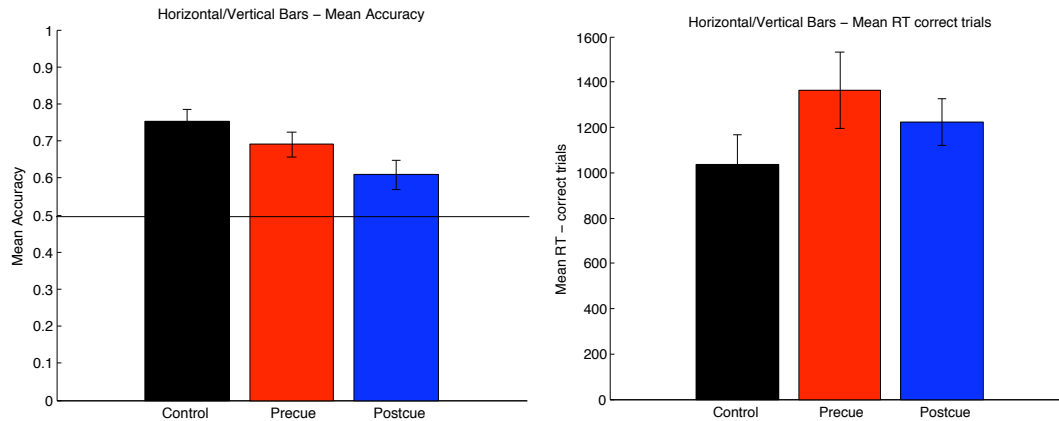
Figure 2.21: Exp. 1.4 Accuracy and Reaction Times.

Accuracy (left) and reaction times (right) of the visual search experiment with horizontal and vertical lines for the three experimental conditions. Chance performance was at 50%. Stimulus presentation time was 50 ms ($n = 5$). Error bars reflect 1 SEM.

from the post-cue condition ($p = .052$). As revealed by a planned comparison, accuracy was significantly lower in the post-cue than in the pre-cue condition ($p = .005$). Reaction times were highest in the pre-cue condition and lowest in the control condition (Fig. 2.21). There was again a main effect of condition (F(2,8) = 5.9, $p = .027$) with only the control condition differing from the pre-cue condition ($p = .032$). All other comparisons were not significant ($p > .05$).

When analysed in more detail, performance in the control condition exhibited no effect of target presence (target absent versus target present trials) nor colour heterogeneity (uni- versus mixed-coloured trials) in neither accuracy nor reaction times. When performance was

analysed as a function of the total number of items in the display, neither accuracy nor reaction times varied significantly, i.e. performance curves were more or less flat.

In the cued conditions, neither accuracy nor reaction times varied in dependence of neither the total number of items, nor the number of the target set, nor the number of the distractor set. [3]

2.5.4   Discussion

As for the control condition, the fact that neither accuracy nor reaction times varied with the total number of items confirms that the current task belongs to the type of parallel or "pre-attentive" feature search as traditionally defined (Treisman & Gelade, 1980). Also in the cued conditions, neither accuracy nor reaction times varied with the total set size nor with the number of items in the distractor or the target subset. Therefore, this also indicates presumably attention-free feature search. Surprisingly, however, cueing resulted in a clear accuracy effect: subjects were significantly better in reporting the odd-ball in a subset that they could direct their attention to compared to when they could not. Even when employing the alternative strategy, i.e. searching for the target and reporting its colour, advance knowledge of which colour to look for seems to pose an accuracy advantage.

Reaction time data is less conclusive as pre- and post-cueing did not produce significant differences despite of a significant main effect when taking the control condition into account. In general, reaction times were higher in the cued than in the control condition, which might be unsurprising given that the control condition did not require any selection by colour.

Given that cueing was hypothesised to be an attentional manipulation, it is interesting that it had such a clear accuracy effect on a presumably "pre-attentive" task exhibiting flat performance slopes. There could be several explanations for this: (1) Performance slopes by

---

[3]The same experiment was carried out using diagonal bars (left- and right tilted), in principle the same stimuli tilted by 45 degrees. Subjects were unable to do the task in the cued conditions, 10 out of 12 subjects performed at or below chance level. However, 10 subjects performed reasonably well in the control condition. In the control condition, accuracy was higher in the target absent case than in the target present case (on average 73% for target absent trials, 59% for target present trials). This effect interacted with the total number of items. There was no effect of total number of items, no effect in reaction times, and no effect of colour heterogeneity.

itself may not be sufficiently informative about the attentional requirements of a task, also not a visual search task (c.f. Di Lollo et al., 2001). (2) Cueing might have a different attentional effect than that measured by performance slopes, for example by tapping into feature-based attention rather than spatial attention. (3) As hypothesised, pre-cueing facilitates the perceptual grouping of targets and/ or the selection of the target set. Search for an orientation odd-ball might still largely occur in parallel, but either grouping and selecting the target set or determining the colour of the detected item (depending on the strategy employed) might be attentionally demanding. However, the luminance differences in the stimuli should be fairly easy to detect and as there are just two alternatives, it is surprising that detection accuracy is affected so strongly by cueing.

After all, an effect of attentional set on the current search task might not be as surprising. As reviewed in Chapter 1, there is good evidence that detection of an orientation odd-out does not occur as "pre-attentively" as previously thought (e.g. Joseph et al., 1998; Di Lollo et al., 2001).

Nakayama & Mackeben (1989) employed a similar task with spatial cueing. Their display was composed of black and white bars and the simple search task required subjects to detect the horizontal target among vertical distractors. Subjects detected not only the presence of the target but also reported its colour. Even at very short stimulus presentation times (as short as 33 ms) accuracy was the same for uncued versus spatially cued target positions. The authors concluded that if the target acts as its own cue (i.e. it "pops out") than spatial cueing does not have an effect. The results from the current experiment suggest that feature-based cueing has a rather differential effect than spatial cueing.

With respect to the enumeration experiments, the results indicate that the cueing manipulation interferes also with a process other than enumeration. Being able to direct attention to the relevant feature also poses an advantage in a supposedly "pre-attentive" visual search task, indicating that the cueing manipulation is likely to have tapped into the perceptual grouping of items and/or the selection of the target set.

2.6   General Discussion

The experimental paradigm introduced above tested whether small numerosity information can be extracted from two subsets in parallel. Based on the initial hypothesis, attentional cueing should not have a large effect if subitizing is a truly pre-attentive task. The results of Experiment 1.1. clearly show, however, that being able to attend to the relevant feature of the subset to enumerate poses a clear advantage over the situation when attention has to be spread to two subsets. This is evidence against a parallel and pre-attentive subitizing mechanism and suggest instead that the retrieval of numerosity from a visual display is not as automatic as previously assumed.

It appears that the extraction of numerosity is contingent on other processes that occur prior to enumeration. The results of Experiment 1.2. showed that if grouping of items into subsets is made difficult, enumeration breaks down. Experiment 1.3. ruled out that detection and discrimination of items are responsible for the failure of enumeration. Together with the results from the enumeration experiment, this suggests that mere detection and identification of items is not sufficient for numerosity information to be extracted in parallel, implying that numerosity information does not come "for free". The fact that additional, possibly attention-demanding processes are involved in enumeration may be an argument against an automatic and effortless retrieval of numerosity information from a visual display: if any of the processes that enumeration depends on requires attention due to the stimulus properties or to the attentional set, then enumeration, including subitizing fails or is impaired. This interpretation is in accord with Trick and Pylyshyn (1993) who in fact were the first to show that the flat reaction time curve for subitizing disappears if target-distractor discrimination requires attention (as it is the case with conjunction target-distractor combinations). Within their FINST theory, Trick and Pylyshyn (1993) attribute the effect of attention on the level of FINST assignment. Here it is suggested that attentional constraints may influence a number of processes that enumeration is contingent on without assuming a limited number of index tokens.

2.6.1   Possible Criticisms

Experiment 1.4. showed that the cueing manipulation might have affected the grouping of items or the selection of a target set. As such it is unclear whether the attentional manipulation of cueing affected only enumeration or the processes enumeration is contingent on. It is conceivable that the extraction of numerosity itself does not require attention once the detection, identification, grouping and set selection can be achieved pre-attentively. However, these processes can often be in themselves attentionally demanding (e.g. Mack et al., 1992; Joseph, Chun & Nakayama, 1997). If subitizing is dependent on attentionally demanding processes, it cannot be a pre-attentive process altogether. The only situation under which subitizing might be a pre-attentive task is when the visual input is impoverished enough to make both detection, individuation and enumeration a very easy task, i.e. in the case where just a few salient dots are displayed well-separated on a uni-coloured background without any distractors present. However, this is a very artificial setting and the validity of such pre-attentive subitizing for real-world vision is questionable. The current paradigm even offered such an impoverished situation (black and white dots on a grey background) and yet two spatially intermixed subsets could not be enumerated in parallel.

The Role of Perceptual Grouping

The question whether perceptual grouping requires attention remains controversial. Using an inattentional blindness paradigm, Mack et al. (1992) showed that perceptual grouping by lightness and proximity does not occur under conditions of inattention. Furthermore, when attention is engaged by a secondary task, grouping by similarity and proximity is impaired (Ben-Av, Sagi & Braun, 1992). However, when measuring grouping indirectly (via its implicit effect on perceiving a visual illusion), Moore and Egeth (1997) and Lamy, Segal and Ruderman (2006) demonstrated that grouping can occur under conditions of inattention. From these latter studies, it appears that asking subjects for an explicit report of what they have seen (as employed in the studies by Mack et al. (1992) and Ben-Av et al. (1992)) may not be a very reliable measurement and grouping may occur without being available for explicit report.

With respect to the current experiment, these findings have two implications. First, perceptual grouping might well have occurred pre-attentively and might have in fact not been the process that was limited by the cueing manipulations. The "indirect measure" of perceptual grouping was enumeration and, if assumed to be a separate task, this might have been the main process being affected by attentional cueing (as originally hypothesised). Second, the findings from the grouping literature point to an alternative explanation. It is conceivable that in the post-cue condition, the numerosity of both subsets was retrieved successfully, but that the results for both subsets were not available for explicit report. This interpretation implies that reporting the numerosity of two subsets instead of one is a more difficult task. This argument is closely related to the one of a possible memory constraint as discussed in the following.

2.6.2   The Role of Memory

As the post-cue condition required subjects to retrieve the numerosity of two subsets instead of one in order to give an equally accurate judgement than in the pre-cue condition, the increase in memory load could have caused the observed effects.

Keeping in mind that subjects had to handle only two subsets and no further distracting stimuli or secondary task, evidence from the literature on visual short-term memory (VSTM) weakens this argument. It is commonly assumed that visual short-term memory has a capacity limit of about 4 items (Luck & Vogel, 1997), although this might in fact vary from 1.6 to about 5 items depending on stimulus complexity (Alvarez & Cavanagh, 2004). In the case of relatively low stimulus complexity, it is safe to assume that the encoding of two features of a visual scene lies within the capacity limit of visual short-term memory. If one supposes that small numerosity information is an automatically extracted feature of a visual scene, comparable to colour or orientation information, then keeping the numerosity from 2 subsets in short-term memory is within this capacity limit. Therefore, the behavioural disadvantage in the post-cue condition cannot be entirely explained by a memory effect under the assumption of an automatic subitizing mechanism. If numerosity information of two subsets can be held in short-term memory, it should also be available for explicit report.

Even under the assumption that enumeration is a possibly memory-demanding process

and thus constrained by the same capacity-limit (about 4 items), cueing effects should not have occurred within the subitizing range as this range coincides with the capacity limit of VSTM. That is, when a subitizable set was the post-cued target set (which was most often the case given the few trials in the counting range), then enumerating this set correctly should not have been constrained by a VSTM capacity limit. The possibility of subitizing as a memory-demanding process and a possible link to visual-short term memory will be elaborated in Chapter 5 (General Discussion).

Furthermore, the timings employed in this experiment should not hinder encoding or retrieval of information from VSTM. In the classical change detection paradigms testing the VSTM capacity limit, stimulus presentation times around 100 ms (Vogel, Woodman & Luck, 2001) and retention intervals of up to 1200 ms (Todd & Marois, 2004) are typically used. As the disadvantage of post-cueing was also observed in the subject group being tested at 150 ms stimulus presentation time, the effect of post-cueing cannot be explained by a failure in memory encoding. Furthermore, the delay between stimulus onset and cue onset was at most 650 ms and as such well within the typical retention period of VSTM (even if accounting for cue processing). Thus, the disadvantage in the post-cue condition cannot be due to simple "forgetting".

Thus, under the assumptions of a pre-attentive subitizing mechanism, the behavioural effect of post-cueing is unlikely to be caused purely by a memory effect. However, if subitizing is attentive, then the observed effects might also have a memory component in addition to an attentional component. Attention plays an important role for information to access visual short-term memory (e.g. Duncan & Humphreys, 1989), so under conditions of distributed attention, encoding into memory or retrieval from memory might be impaired.

The Role of Feature Binding

There is another alternative explanation on how the disadvantage of the post-cue conditions might have come about. It is conceivable that numerosity is retrieved successfully for both sets but the binding of numerosity and colour failed. In fact, the results of the visual search experiment (Exp. 1.4) indicate that post-cueing may affect the binding of colour with the detected target. In the case of enumeration, subjects might have retrieved the numerosity correctly but

mistakenly reported the numerosity of the distractor set. This possibility was explored (Silvia Pagano, personal communication) by analysing whether in incorrect trials, subjects reported the distractor numerosity more frequently than other numerosities. This was not the case; however, incorrect trials are noisy data and as such difficult to interpret. Given that feature binding is commonly assumed to require focal attention (Treisman & Gelade, 1980), this criticism is a valid argument. It should be noted though that in the post-cue condition, attention was not absent, but distributed. It might be an interesting conclusion from these experiments, particularly from Experiment 1.4, that distributing attention to two features instead of one has an effect on feature binding.

2.6.3   Conclusions

This series of experiments showed that subitizing is affected by manipulations of attention such that accuracy is decreased when attention has to be distributed to two sets of items. Small numerosity information can therefore not be extracted in parallel from two subsets with equal accuracy than from one subset. This is evidence against an automatic and pre-attentive subitizing mechanism.

This study furthermore suggests that subitizing depends on perceptual grouping. Thus, any manipulation that interferes with grouping makes subitizing impossible or less accurate. The employed paradigm left it unclear, however, whether the employed attentional manipulations interfered with the binding of numerosity with a certain subset or whether they interfered with enumeration *per se*.

Chapter 3

Experiment 2: Investigating Subitizing under Attentional Load: A Dual Task Paradigm

3.1   Introduction

This experiment established another approach to test the hypothesis of a parallel and pre-attentive subitizing mechanism. Given that the experimental manipulation of Experiment 1 was slightly confounded with respect to whether it was purely attention or also memory that was manipulated by cueing, this experiment attempted to circumvent this confound by systematically varying the amount of attentional resources that could be allocated to an enumeration task.

As mentioned in the General Introduction, the definition of a pre-attentive process entails that it should be unaffected by how much attentional resources can be allocated to the task, supposing that processing happens on a level *prior* to any involvement of attention. One way of testing pre-attentive processing therefore is to impose an additional, attention-demanding distractor task and to measure the effect on that task that supposedly involves pre-attentive processing.

Such an approach has been applied in a pioneering study by Joseph, Chun and Nakayama (1997). They tested whether a simple feature detection task belonging to the classical types of "pre-attentive" tasks is affected by a competing, attention-demanding task. The primary task (which is always defined as the attention-manipulating task) consisted of the detection of a target letter in a rapid serial visual presentation (RSVP) of distractor letters at fixation. The secondary task (which is always defined as the task to be tested for attentional dependence) consisted of the detection of a salient orientation odd-ball in a circular array of Gabor patches arranged around central fixation. The results strikingly demonstrated that when the target letter was displayed at the same time as the ring of Gabor patches (zero SOA), odd-ball detection accuracy was reduced to 60%, compared to single task performance at more than 90%. The bigger the lag between the appearance of the target letter and the array of Gabor patches, the better odd-ball detection became and performance was restored at a delay of about 700 ms.[1] Thus, even a simple feature detection task, previously defined as pre-attentive (and as such identified in control experiments by means of a flat search function) was severely impaired when attention was diverted to an additional task. It is apparent why these experiments had a considerable impact in the field of cognition: they raised the question whether there is *any* perceptual judgement that can be done without provision of attentional resources and whether the idea of "pre-attentive" processing might be a conceptual artefact altogether.

A general criticism on dual task paradigms can be made, however. It is unsurprising that in the case when subjects have to distribute their attention to two tasks instead of one, performance will drop in both tasks compared to single task performance. A solution to this problem is offered by Lavie's load theory of attention (Lavie, 1995; 2005). Load theory states that the processing of irrelevant distractor information (typically that of a secondary task) crucially depends on the attentional requirements, i.e. the load of the target information (typically that of a primary task). Assuming limited capacity for perception, it thus depends on the distribution of such capacities to what extent irrelevant information is processed. If processing target information (e.g. a primary task) is of high load, it takes up the majority of processing capacity and does not leave much resources for distractor processing (e.g. a secondary task). Under these circumstances, distractor processing is reduced or sometimes even eliminated. If,

---

[1]The inability to redistribute attention shortly after the detection of a target to a subsequent event in an RSV presentation is also termed "attentional blink" (Raymond, Shapiro & Arnell, 1992).

however, the attentional demands of target processing require only little resources and thus being of low load, capacity involuntary "spills over" to the processing of distractors.

A consequence of this approach is that the crucial comparison is not between single and dual task conditions (thus whether attention is distributed or not) but, within dual task conditions, between those conditions when the primary task is highly attention-demanding (high load) and those when it is not as highly attention-demanding (low load). Thus, in both low and high load condition, attention needs to be distributed, but the balance of the distribution changes. Typically, the stimuli for the primary task are chosen so that they are the same in low and high load, but task requirements are specified so that under low load, the attentional demands are little whereas under high load, they are greater. Thus the perceptual complexity between load and high load conditions does not differ, rather it is the extent to which attention needs to be allocated to make the required judgement. By varying the load of the primary task, this approach offers the possibility to manipulate attentional resources allocated to a secondary task in a more balanced way than a standard dual task paradigm.

3.1.1   The Paradigm

In order to test the hypothesis of a pre-attentive subitizing mechanism, visual enumeration was tested under conditions of dual task situations and variable attentional load. The paradigm for the current experiment was inspired by Joseph, Chun and Nakayama (1997) and was adapted to implement the approach from load theory (Lavie, 1995; 2005). A dual paradigm was designed, composed of a colour detection task at fovea as primary task and a more peripheral enumeration task as secondary task. The stimulus for the primary task consisted of a bi-coloured diamond shape (see example stimulus in Fig. 3.1) that was concurrently displayed with a ring of Gabor-like patches for the secondary enumeration task. I.e. in contrast to Joseph, Chun and Nakayama (1997), no rapid serial visual presentation was employed. The primary task implemented the manipulation of attentional load: under low load, subjects were instructed to detect a simple feature (the colour red) whereas under high load they were asked to detect a specific feature conjunction (a colour-orientation conjunction). Many visual search studies have demonstrated that feature detection is much less attention demanding than conjunction detection (e.g. Treisman

& Gelade, 1980) and both have been proven to be effective in manipulating load (Lavie, 1995). This specific colour detection task was established as it was hypothesised not to involve any numerosity or magnitude related processing. Thus, a potential interference with the secondary enumeration task unrelated to purely attentional effects was avoided.

The stimulus of the secondary task was similar to the one used by Joseph, Chun and Nakayama (1997): a ring of Gabor-like patches arranged around the central colour target. In Experiment 2.1, 1 up to a maximum of 8 targets were displayed amongst distractors and subjects reported the number of targets in the ring. Equiluminant Gabor-like patches of high and low contrast were used (as targets and distractors, respectively) so that subjects could not make their judgement based on luminance but had to rely on numerosity. As in Experiment 1, the total number of items as well as the number of distractors was variable for each target number so that subject's numerosity judgement could not be facilitated by those covarying parameters. Again, both enumeration within the subitizing range as well as the counting range was tested in order to study the effects of dual task and attentional load across both ranges. Note again that verbal counting was suppressed by employing short stimulus presentation times and the term "estimating" is more appropriate for the enumeration of 5 to 8 items.

Two experiments were carried out. Experiment 2.1 investigated the effects of dual task situations and attentional load on enumeration, whereas Experiment 2.2 investigated the same effects on single target detection.[2]

3.2   Experiment 2.1 Effects of Dual Task and Attentional Load on Enumeration

3.2.1   Rationale

This experiment explored how imposing an additional task and varying its attentional demands affects visual enumeration. It was hypothesised that if subitizing is a truly pre-attentive task,

---

[2]Experiment 2.1 is published as a shortened version in Vetter, Butterworth & Bahrami (2008). Modulating attentional load affects numerosity estimation: Evidence against a pre-attentive subitizing mechanism. *PLoS One*,**3(9)**, e3269.

it should be unaffected by the manipulation of dual task and attentional load. The definition of "pre-attentive" implies that processing occurs prior to any allocation of attention, i.e. having to allocate some attentional resources to an additional task as well as the attentional load should not matter with respect to the outcome of the task. That is, subitizing accuracy should be the same in single task and low load conditions, and, more importantly, should not differ when comparing low load with high load conditions. However, if subitizing is an attentive task, then subitizing accuracy should be decreased due to both dual task conditions and increasing attentional load. Estimation of numerosities 5 to 8 was predicted to be attentive and thus should also be affected by both experimental manipulations.

3.2.2   Methods

Subjects

14 subjects (mean age: 23.1, range: 18-30, 10 females) with normal or corrected-to-normal vision were tested. All gave written informed consent and were paid for their participation. Subjects sat in a dimly lit room and viewed the stimuli at a distance of 50 cm from the computer screen using a chin rest.

Stimuli

The visual stimulus consisted of: (1) a central diamond shape comprising four coloured triangles and (2) a circle of Gabor-like patches on a grey background (see example stimulus in Fig. 3.1). The employed Gabor patches were cut off at the edges (thus forming sharp-edged round stimuli containing a Gabor grating) to create stimuli that are clearly segregated from the background. Eight different colour combinations were used for the central diamond shape. Each colour combination was defined either as target or non-target for either low or high load condition. The Gabor-like patches in the circle were either vertically orientated high-contrast (100%) targets or horizontally orientated low-contrast (50%) distractors. The positions of high amongst low contrast patches were randomly assigned for each trial. Patches were always equally spaced from each other and placed in the circle with a random offset from 0° clockwise to prevent occupation
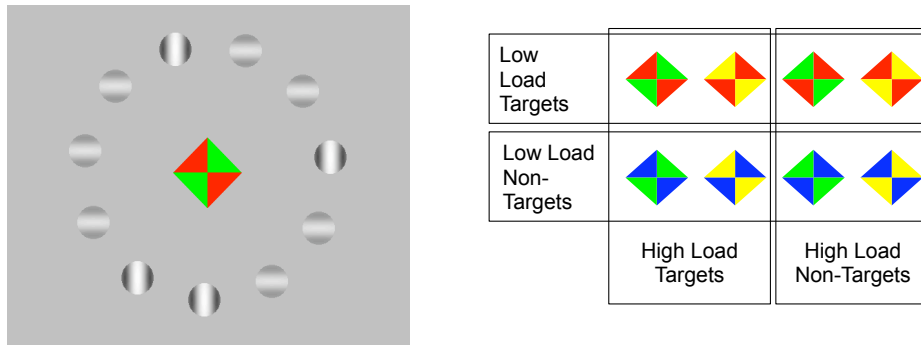
Figure 3.1: Stimulus Example and Colour Combinations of Exp. 2.1

Left: Stimulus Example of Exp. 2.1. The primary task was a colour detection task at
fovea, the secondary task required to enumerate the high contrast Gabor-like patches
in the ring. Right: Colour combinations used for low and high load in the primary
task.

of the same positions in each trial. The whole circular stimulus was 16° wide, with the central

diamond shape comprising 4° and each patch 2°. The grey value of the background was adjusted

to mid-grey and gamma corrected for output luminance (as was the Gabor value). Stimuli were

generated using the Cogent toolbox (www.vislab.ucl.ac.uk/Cogent/) for MATLAB (Mathworks,

Inc).

Task and Experimental Procedure

A dual task paradigm was employed. The primary task was a speeded target detection task at

fovea which implemented the manipulation of attentional load. Under low load, subjects were

asked to detect a simple feature in the central diamond shape (the colour red, regardless of spatial arrangement). Under high load, subjects detected specific conjunctions of colour and spatial arrangement: either two green triangles aligned along the right-tilted diagonal or two yellow triangles aligned along the left-tilted diagonal. Importantly, subjects were instructed not to respond to the opposite combinations (see Fig. 3.1 for colour combinations of low load and high load targets, respectively). The colour combinations were chosen so that targets and non-targets were equally distributed in each load condition. Both low and high load condition therefore consisted of the same set of stimuli, only the task instructions changed.

As a secondary task, subjects judged the number of the vertically oriented high-contrast Gabor-like patches (targets). The horizontally oriented low-contrast patches served as distractors. The number of targets ranged from 1 to 8, total number of items in the circle ranged from 9 to 13, counterbalanced for each target number and load condition. Distractors were used to de-correlate task difficulty from the overall processing effort required for multiple stimuli. The number of distractors did not co-vary with the number of targets. Therefore, numerosity judgement could be made neither on the basis of the total number of items present nor on the basis of the number of distractors. As distractors were equally luminous than targets, numerosity could not be judged based on overall luminance either. The only parameter that co-varied with numerosity was contrast.

Within each trial, after a fixation cross of 1 second, the stimulus was displayed for 200 ms and was followed by a mask. The mask stayed on the screen until subjects responded (see Fig. 3.2). Inter-trial intervals varied randomly between 1 and 2 seconds. Note that verbal counting was prevented by employing short stimulus durations.

Subjects always responded first to the primary task and subsequently to the secondary task, ensuring that attentional resources were manipulated by the processing requirements of the primary task and not by the number of items in the secondary task. Subjects responded with their right hand on two adjacent keys to the primary task and subsequently, as accurately as possible, with their left hand to the secondary task using number keys 1 - 8. Overall, accuracy was emphasised over speed. Subjects were given practice trials before each single task and dual task condition and had the opportunity to take a break after each block. The testing session
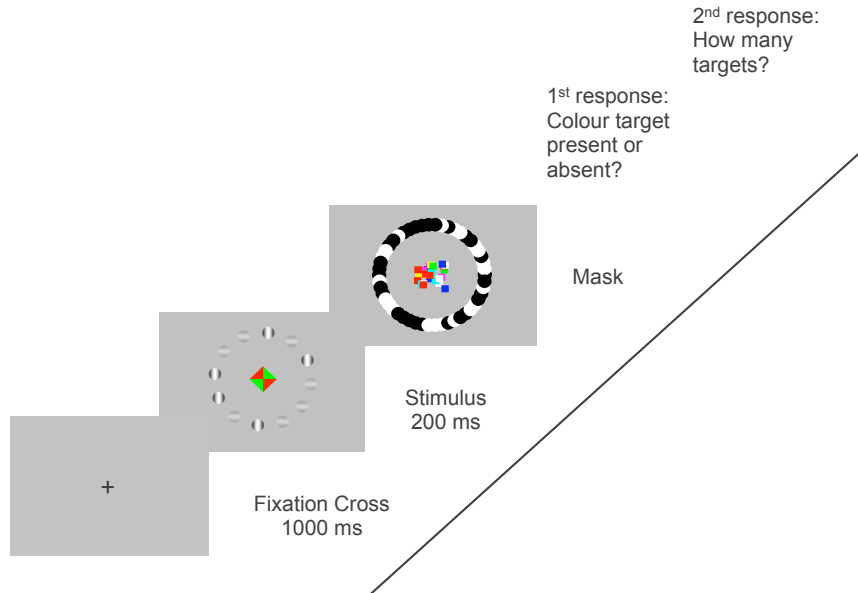
Figure 3.2: Experimental Procedure of Exp. 2.1

Under dual task conditions, subjects responded both to the central colour detection task as well as to the numerical task. Under single task conditions, only one response was required.

lasted for about one hour.

Experimental Design

Each colour combination of the primary task was combined once with each numerosity target of the secondary task, resulting in 64 trials per block. Subjects first performed 2 blocks of each task under single task condition. That is, the numerical task was performed ignoring the central colour shape and the target detection task was performed ignoring the Gabor-like patches (1 block low load, 1 block high load). Subjects were therefore well trained in each of the two tasks before being tested under dual task conditions. 4 blocks of dual task were performed, 2 blocks

of low load and 2 blocks of high load in the order ABBA or BAAB, counterbalanced across subjects. Each subject performed 16 trials per target number per experimental condition (512 trials for the whole experiment).

Data Analysis

Reaction time and accuracy data of the primary task were compared using a repeated measures ANOVA with within-subject factors load condition (2 levels: low load and high load) and task (2 levels: single task and dual task). Accuracy, reaction times, mean responses, mean response variance and Weber fraction of the secondary task were analysed with a repeated measures ANOVA with within-subject factors experimental condition (3 levels: single task, low load and high load) and target number (either 8 levels for the whole number range, or 4 levels separately for the subitizing range (1-4) and the estimation range (5-8)). Post-hoc comparisons between experimental conditions were corrected for multiple comparisons after Bonferroni. Mean responses were analysed for deviation from the correct response using one-sample T-tests. The same tests were applied to determine whether accuracy was above chance level. Given that trial frequencies were the same at each numerosity level, chance performance was defined as an eighth (.125).

3.2.3   Results

Primary Colour Detection Task

Accuracy and reaction time data of the central colour detection task under both low and high attentional load are plotted in Fig. 3.3. As predicted, subjects responded more slowly under high attentional load compared to low attentional load ($F(1,13) = 114.57$, $p < .001$) and significantly less accurately ($F(1,13) = 20.26$, $p = .001$). Subjects were also slower under dual-task conditions ($F(1,13) = 97.77$, $p < .001$) and less accurate ($F(1,13) = 37.01$, $p < .001$) compared to single task conditions. There was a significant interaction between task and condition for accuracy ($F(1,13) = 17.80$, $p = .001$) but not for reaction time ($F(1,13) = .133$, $p > .05$). These results confirm that the manipulation of attentional load was effective.

129
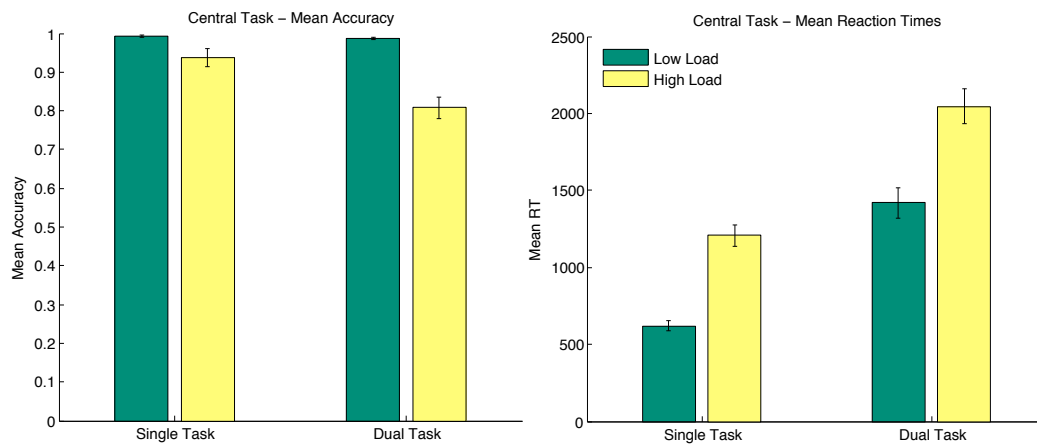
Figure 3.3: Exp. 2.1 Accurcy and RT - Central Colour Task.

Results of Exp. 2.1 for the central colour detection task. Mean accuracy (left) and mean reaction times (right) are plotted for both low and high load condition. Error bars indicate 1 SEM.

Secondary Numerical Task

Accuracy   Accuracy data for the secondary numerical task are depicted in Fig. 3.4 and statistical results are listed in Table 3.1 for the whole numerosity range as well as for the subitizing and the estimation range separately. Overall, enumeration accuracy declined steadily with increasing numerosity forming a sigmoidal performance curve. Performance was best under single task conditions and decreased with increasing attentional load. Under single task conditions, subjects performed at chance level (.125) at numerosities 7 and 8 (t $\leq$ 1.18, p > .05), whereas under both dual task conditions, accuracy reached chance already at numerosity 6 (t $\leq$ 1.53, p > .05).

Dual task and attentional load had a clear deteriorating effect on enumeration performance. Accuracy dropped significantly between single task and low load conditions, and more importantly, performance under high load was even more impaired than under low load. The effect was particularly pronounced in the subitizing range. In the estimation range, accuracy was low overall and differences between experimental conditions were less pronounced. Nevertheless, there was a main effect of condition, mainly due to a significant difference between single task and high load condition. As performance reached chance level for numerosities 7 and 8, the analysis was repeated for the numerosity range 5 to 6. There was still a main effect of condition ($F(2,26)$ = 3.55, p = .043) but post-hoc comparisons did not reach significance.

As expected, accuracy decreased with increasing target number, apparent as a main effect of target number in the whole numerosity range as well as in both subitizing and estimation range. Target number interacted with condition only when taking the whole numerosity range into account, but this interaction did not reach significance when analysing both sub-ranges separately.

Reaction Times   Due to the sequential key responses, reaction time data are not very meaningful in this paradigm. Subjects were asked to process both primary task and secondary task stimuli simultaneously, but responded always first to the primary task and subsequently to the secondary task. Thus, with increasing load, responses were slowed due to the processing demands of both the primary and the secondary task and it is impossible to infer the extent to
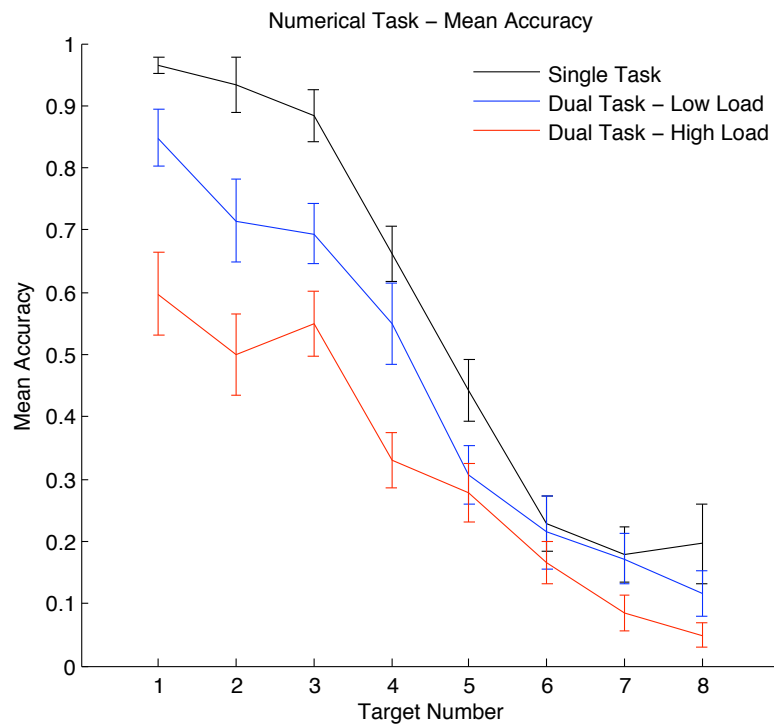
Figure 3.4: Exp. 2.1 Numerosity Task - Accuracy.

Accuracy data (proportion correct) of the numerosity task of Exp. 2.1 as a function of target number for each load condition. Error bars indicate 1 SEM.

Exp. 2.1 Accuracy - Numerical Task

		Numerosity Range		
		1 - 8	**1 - 4**	**5 - 8**
Within-Subject Effects				
Condition	p	<.001	<.001	.010
	F(df)	42.49 (2,26)	58.88 (2,26)	5.55 (2,26)
Target Number	p	<.001	<.001	<.001
	F(df)	92.65 (7,91)	15.94 (3,39)	32.73 (3,39)
Condition x Target Number	p	<.001	n. s.	n. s.
	F(df)	5.62 (14,182)	1.29 (6,78)	1.59 (6,78)
Post-hoc Comparisons				
Condition Single Task vs Low Load	p	.004	.003	n. s.
Single Task vs High Load	p	<.001	<.001	.007
Low Load vs High Load	p	.002	<.001	n. s.

Table 3.1: Exp. 2.1 Accuracy - Numerical Task

Repeated Measures ANOVA results for accuracy data of the secondary numerosity task of Exp. 2.1. $n = 14$. Non-significant effects indicate $p > .05$. p values in post-hoc comparisons are Bonferroni-corrected.

which each of the two tasks contributed to increased reaction times. In fact, the dual task costs in the primary task suggest that primary and secondary task processing interacted. Nevertheless, reaction time data was analysed to ensure that the accuracy effect in the numerical task was not due to a speed-accuracy trade-off. As expected, reaction times (for correct trials) were lowest in the single task condition (across all numerosities on average 1.53 s), significantly higher in the low load condition (2.10 s), and again significantly higher in the high load condition (2.91 s; main effect: $F(2,26) = 138.39$, $p < .001$, all post-hoc comparisons: $p < .001$). Reaction times also increased with target number ($F(7,91) = 9.59$, $p < .001$), an effect which interacted with experimental condition ($F(14,182) = 2.67$, $p = .001$). Thus, although it is unclear which task contributed primarily to increased reaction times, it is unlikely that the effect of dual task and attentional load on accuracy was entirely due to a speed-accuracy trade-off.

Mean Responses As in Experiment 1, mean responses were analysed as a measure of deviation from the correct response. Mean responses are plotted in Fig. 3.5. At numerosity 1, subjects on average overestimated in all experimental conditions (single task: t = 2.48, $p = .028$, low load: t = 3.51, $p = .004$, high load: t = 4.71, $p < .001$). At numerosity 2, subjects only overestimated in the high load condition (t = 3.23, $p = .007$). Responses did not depart significantly from correct response at target number 3, but underestimation occurred from numerosity 4 onwards (t $\leq$ -2.7, $p \leq$ .018 in all comparisons for all numerosities 4 - 8 in all experimental conditions). Analysis of mean response deviations showed that underestimation in this range increased with numerosity ($F(4,52) = 115.93$, $p < .001$).

There was a significant main effect of experimental condition ($F(2,26) = 9.59$, $p = .001$), due to a difference between the single task condition and the high load condition (post-hoc comparison: $F(2,12) = 8.55$, $p = .003$, all other post-hoc comparisons $p > .05$). Naturally, there was a significant effect of target number ($F(7,91) = 231.09$, $p < .001$) that interacted with condition ($F(14,182) = 26.67$, $p < .001$), i.e. overestimation in the low numerosities and underestimation in the high numerosities occurred stronger in the high load condition than in all other conditions.
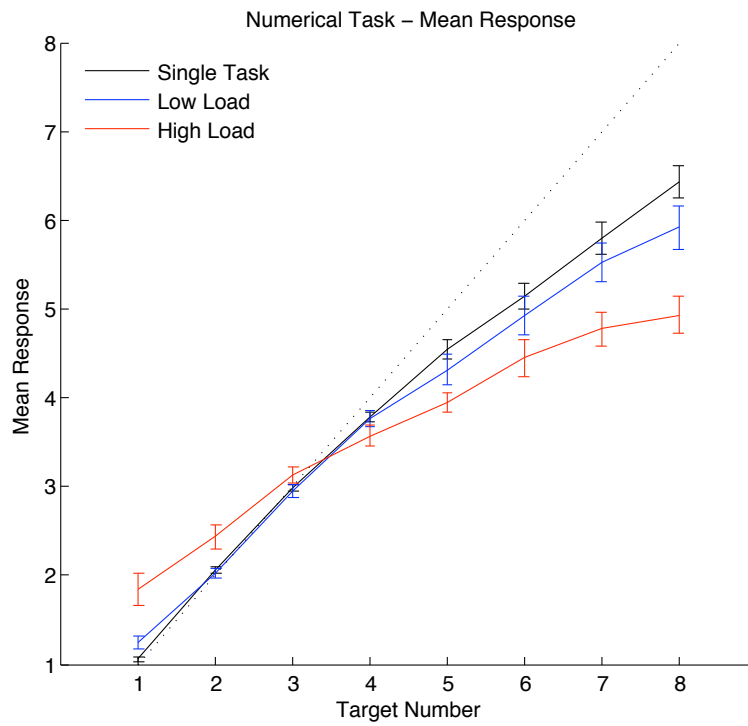
Figure 3.5: Exp. 2.1 Mean Response - Numerosity Task

Mean responses of Exp. 2.1 as a function of target number. The diagonal line indicates perfect performance. Error bars reflect 1 SEM.
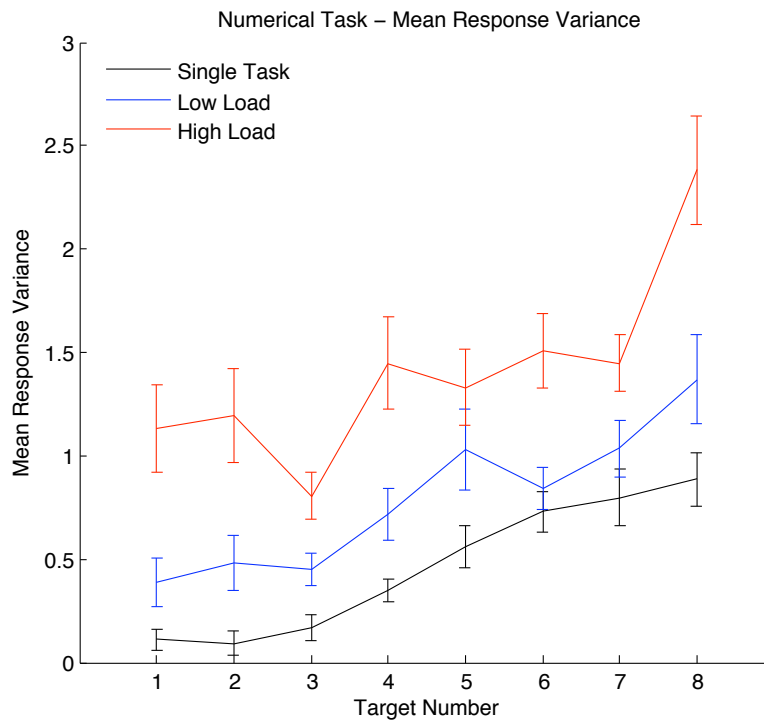
Figure 3.6: Exp. 2.1 Mean Response Variance - Numerosity Task

Mean response variance of Exp. 2.1 as a function of target number. Error bars reflect 1 SEM.

Mean Response Variance   The variance of mean responses was analysed as a measure of subjects' response variability. Results are depicted in Fig. 3.6. Response variance increased significantly from single task to low load to high load conditions (main effect: $F_{(2,26)} = 75.14$, $p < .001$, post-hoc comparisons: single task versus low load $p = .001$, single task versus high load $p < .001$, low load versus high load $p < .001$). Variances also increased with numerosity ($F_{(7,91)} = 15.91$, $p < .001$) and this effect interacted with the effect of condition ($F_{(14,182)} = 2.60$, $p = .002$). All effects replicated when conducting the analysis separately for the subitizing and estimation range.

Weber Fraction   As a measure of discriminability, Weber fraction (response standard deviation divided by target number) was analysed (Fig. 3.7). The higher the Weber fraction, the

lower the discriminability (the bigger the necessary difference between stimulus size so that it can be detected successfully). Weber fraction was consistently higher under dual than under single task conditions, and again higher under high than under low attentional load (main effect: $F(2,26) = 84.48$, $p < .001$; all post-hoc comparisons $p < .001$). All effects replicated when subitizing and estimation ranges were analysed separately (all $p < .001$). There was a significant main effect of target number ($F(7,91) = 20.37$, $p < .001$). Results from each experimental conditions were tested separately for an effect of target number to see whether Weber fraction varied with numerosity. In the single task condition, Weber fraction did not differ across numerosities ($F(7,91) = 1.46$, $p > .05$). Under dual task conditions, however, Weber fraction was highest in the low numerosities and decreased towards higher numerosities ($F(7,91) = 30.14$, $p <.001$). Considering all three conditions, the effect of target number interacted with condition ($F(14,182)$ 27.01, $p < .001$).

3.2.4 Discussion

In this experiment, visual enumeration was tested under conditions of constrained availability of attentional resources. It was predicted that if subitizing is a truly pre-attentive process, it should be unaffected when an additional, attention-demanding task is imposed, and crucially, it should not be affected by the attentional load of this additional task. The current results clearly fail to support this prediction. Subitizing accuracy was impaired under dual-task conditions compared to single task conditions, even if the additional task comprised only the detection of a single salient feature (the colour red). More crucially, however, subitizing was even more severely impaired when the additional task required a judgement of high attentional load (a conjunction detection). Thus, the more attentional processing resources were taken away from the numerosity judgement task, the more subitizing ability deteriorated. These results challenge the traditional notion of a pre-attentive subitizing mechanism.

As this experiment was designed to investigate the immediate apprehension of numerosity, short stimulus presentation times were employed and as a result, accuracy was low in the estimation range. Mean responses and response variance, however, demonstrated a clear effect of dual task and attentional load also in the estimation range. The more attentional resources
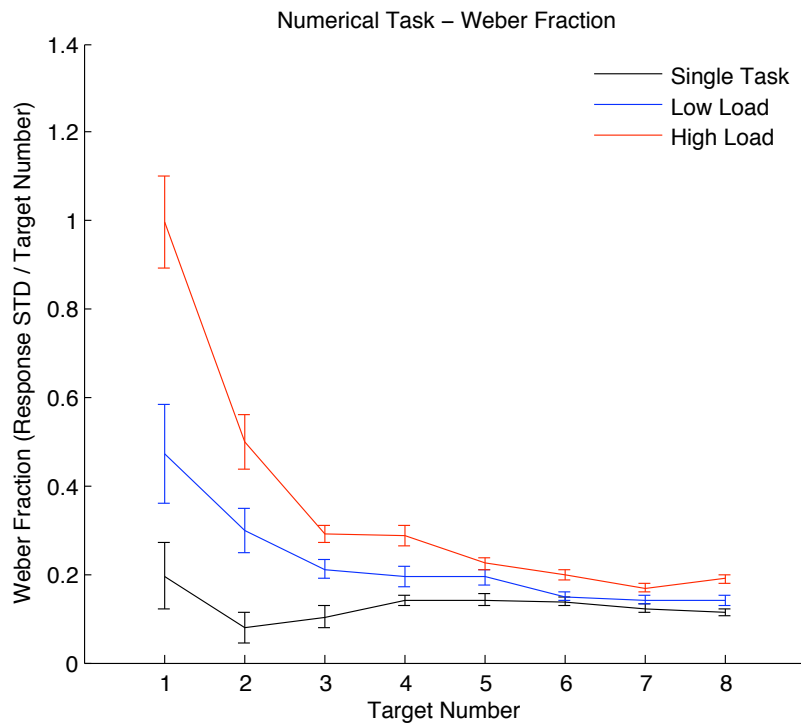
Figure 3.7: Exp. 2.1 Weber Fraction - Numerosity Task

Weber fraction (response standard deviation / target number) of Exp. 2.1 as a function of target number. Error bars reflect 1 SEM.

were drawn away from the enumeration task, the more subjects underestimated and the more variable their responses became. From these results, it appears that numerosity judgement was affected in a systematic manner: the fewer processing resources can be allocated to the task and the more difficult numerosity judgement becomes at higher numerosities, the more performance deviates from an unaffected distribution.

Weber fractions strikingly mirrored the accuracy data and showed that increased attentional load impaired discrimination ability particularly in the subitizing range. Although the difference between experimental conditions was not as large in the estimation range, a clear and significant decrease of discriminability with increasing attentional load was also observed there. Whereas Weber fractions decreased with numerosity in both load conditions, they did not differ across numerosities in the single task condition. This is consistent with the findings of Ross (2003) who demonstrated constant Weber fractions across a wide range of large numerosities. That is, under single task conditions, discriminability is the same across all numerosities. It is, however, affected by attentional load. Again, this result demonstrates that attentional load had a systematic effect on numerosity judgement: identifying the numerosity of a set and discriminating it from the rest of the response possibilities became increasingly difficult with increasing attentional load of the primary task.

The fact that both subitizing and estimation were affected by the manipulation of attentional resources could be interpreted as evidence against a functional dichotomy between these two processes. The present results are consistent with the idea that both subitizing and estimation reflect stages on a single, continuous numerosity judgement mechanism. However, this experiment was not designed to investigate the functional dichotomy of subitizing and counting, and more specific studies are needed to address this issue. Nevertheless, these results render one of the main arguments for such a dichotomy invalid: that subitizing is parallel and pre-attentive and might therefore be different from an attentive counting or estimation stage. In the absence of this argument, the question is raised what makes subitizing different from estimation apart from the obvious difference in processing effort due to the amount of information to be dealt with.

As predicted, the results from the primary colour detection task confirmed that the

manipulation of attentional load was effective, i.e. the high load task was more difficult than the low load task. There was a cost of dual task conditions in this primary task, i.e. as predicted from the criticism of dual task paradigms, performing two tasks instead of one affects the performance in both tasks.

With respect to Experiment 1, it is interesting to note that dual task conditions and attentional load systematically increased response variances, i.e subjects became more variable in their response behaviour and tended to choose numerosities rather distant from the actual target number. The cueing manipulation of Experiment 1, however, did not lead to increased response variance, i.e. subjects were similarly variable in choosing a response in both control, pre-cue and post-cue conditions. This suggests that the experimental manipulations of these two experiments interacted on different levels of numerosity processing. For example, it could be argued that the cueing manipulation interfered with perceptual grouping or feature binding, whereas dual task and attentional load might have interfered with the extraction of numerosity and the ability to discriminate several adjacent numerosities from each other.

Possible Criticism

Given that increased attentional load seems to introduce noise in the response behaviour, it could be argued that the load manipulation might have affected the early perceptual stages of detecting a high contrast Gabor-like patch rather than the later stages of numerosity retrieval from the detected targets. Since enumeration is dependent on detection, the results of this experiment cannot distinguish which level of processing was affected by the employed experimental manipulations. Therefore, a single target detection version of the experiment was carried out in Experiment 2.2 to clarify this issue.

3.3   Experiment 2.2 Effects of Dual Task and Attentional Load on Single Target Detection

3.3.1   Rationale

In order to address the possible criticism that low-level visual processing related to the detection of targets was responsible for the effect observed in Exp. 2.1, a single target detection version of the secondary task was designed. Instead of several high-contrast Gabor-like patches, only one or none was present, and subjects decided upon the target's presence or absence (equivalent to a pop-out visual search task). It was hypothesised that detecting a single high-contrast patch among low-contrast patches can be regarded as the most "pre-attentive" version of the experiment. The primary task remained unchanged.

It was predicted that if the primary task affected mainly the detection (and/or discrimination) of high contrast Gabor-like patches, a load-induced drop in accuracy for the target present trials should be observed, in particular with a similar effect size than the one observed for numerosity 1 in Experiment 2.1. However, if the primary task affected numerosity judgement more than target detection, it was predicted that a load-induced effect on the target detection task should either be much smaller than judging numerosity 1 or absent. However, given that Joseph, Chun and Nakayama (1997) found a decrease in accuracy for the detection of a single orientation odd-ball when introducing an attention-demanding secondary task, it would not be entirely surprising to find an effect of attentional load on high contrast target detection as well.

3.3.2   Methods

Subjects

11 subjects were tested (mean age: 22.6, range: 20-30, 10 females) under the same conditions as in Exp. 2.1.

Stimuli

The central colour stimulus was identical to Experiment 2.1. In the periphery, either one high contrast Gabor-like patch was presented amongst low contrast distractors (50% of the trials) or all patches were of low contrast. Total number of items varied pseudo-randomly between 9 to 13 and was counterbalanced for load conditions and target present/ target absent trials. All other aspects of the stimulus were identical to Experiment 2.1.

Task and Experimental Procedure

Instead of judging numerosity, subjects decided whether the single high contrast patch was present or absent in the ring of Gabor-like patches. Subjects always responded first to the primary task with their right hand on two adjacent keys on the keyboard and then to the secondary task with their left hand on two different adjacent keys. All other experimental procedures were the same as in Experiment 2.1. The testing session lasted about 30 minutes.

Experimental Design

Each colour combination of the primary task was combined with 4 trials of the secondary task (2 trials target present, 2 trials target absent), resulting in 32 trials per block. Subjects performed 2 blocks of each task under single task conditions followed by 4 blocks under dual task conditions (2 low and 2 high load blocks in counterbalanced order), resulting in 64 trials per subject per experimental condition (32 trials target present, 32 trials target absent for the secondary task). Total number of trials was 256.

Data Analysis

Accuracy data of the secondary task was analysed with a repeated measure ANOVA with within subject factor experimental condition (3 levels), both for all trials and for target present and target absent trials separately. When comparing detection versus enumeration, accuracy data from the target present trials was combined with accuracy data from numerosity one (resulting

in an additional between-subject factor task).

3.3.3 Results

2 out of 11 subjects performed at chance under high load conditions and were excluded from the analysis.

Primary Colour Detection Task

Accuracy and reaction times to the primary colour detection task are plotted in Fig. 3.8. The results found in Experiment 2.1 were replicated. Subjects responded both less accurately ($F(1,8)$ = 12.13, p = .008) and more slowly ($F(1,8)$ = 101.5, $p < .001$) under high load compared to low load. There was again a cost of dual task versus single task both in accuracy ($F(1,8)$ = 8.26, p = .021) as well as in reaction time ($F(1,8)$ = 68.49, $p < .001$). Task and load interacted in both cases (accuracy: $F(1,8)$ = 9.51, p = .015; RT: $F(1,8)$ = 6.92, p = .030). Thus, the manipulation of attentional load was again effective in this version of the task.

Secondary Single Target Detection Task

Accuracy Accuracy decreased with dual task conditions and with increasing attentional load (Fig. 3.9). There was a significant main effect of experimental condition ($F(2,16)$ = 17.74, $p <$ .001) with post-hoc comparisons showing significant effects between all experimental conditions ($F(2,7)$ = 10.60, single task versus low load: p = .032, single task versus high load: p = .004, low load versus high load: p = .020). In the target present trials, the main effect of condition remained ($F(2,16)$ = 6.07, p = .011) but post-hoc comparisons did not reach significance (all p > .05). In the target absent trials, there was again a main effect ($F(2,16)$ = 8.99, p = .002), but mainly due to the difference between single task and high load condition (p = .018, all other post-hoc comparisons: p > .05).
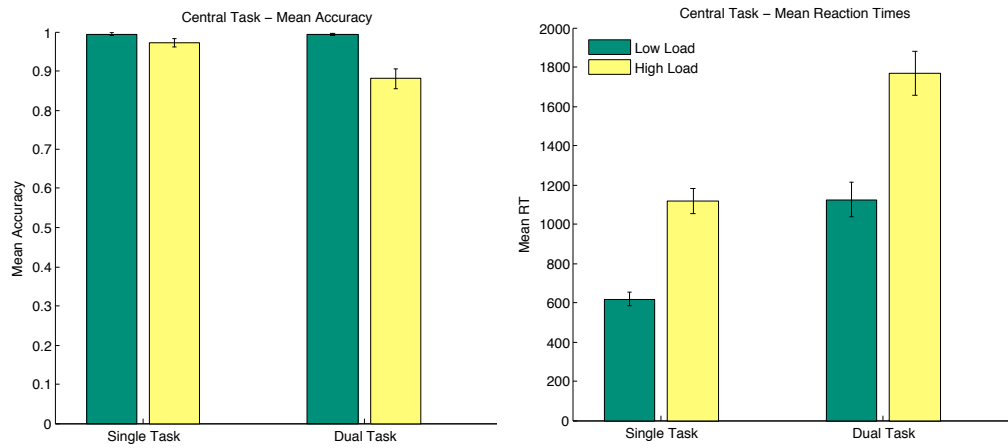
Figure 3.8: Exp. 2.2 Accuracy and RT - Central Colour Task.

Results of Exp. 2.2 for the central colour detection task. Mean accuracy (left) and mean reaction times (right) are plotted for both low and high load condition. Error bars indicate 1 SEM.
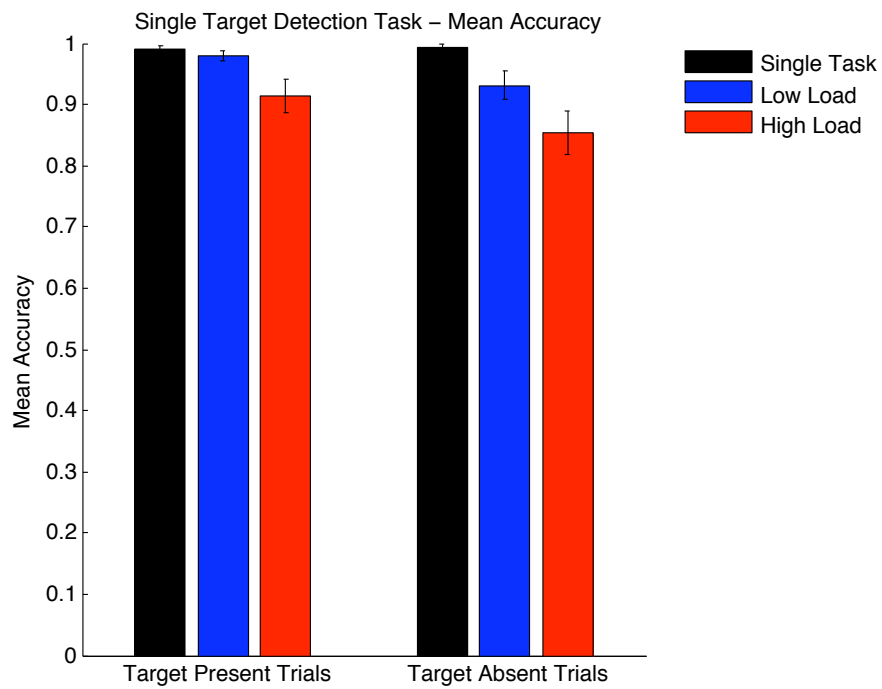
Figure 3.9: Exp. 2.2 Accuracy - Single Target Detection Task.

Accuracy results of Exp. 2.2 of the secondary single target detection task separately for target present and target absent trials for each load condition. Error bars indicate 1 SEM.

Comparing Enumeration versus Detection

Here, enumeration accuracy of numerosity 1 of Exp. 2.1 was compared with detection accuracy in the target present trials of Exp. 2.2. Note that the stimulus was exactly the same in these two situations, only the task instructions and the response possibilities differed. In target detection, high load compared to low load led to an accuracy drop of 6.6%, whereas in enumerating numerosity 1, accuracy decreased by 25.1%. Dual task conditions had a similarly differential effect, low load dual task compared to single task made accuracy drop 1% in detecting an item, but 11.6% in enumerating an item. There was a significant effect of task (detection versus enumeration, $F(1,21) = 10.36$, $p = .004$) as well as a main effect of condition ($F(2,42) = 26.77$, $p < .001$) and a significant interaction (condition x task: $F(2,42) = 11.09$, $p < .001$). Post-hoc t-tests showed that enumeration accuracy differed significantly from detection accuracy under high load ($t = -4.18$, $p = .001$) as well as under low load ($t = -2.76$, $p = .016$). Thus, dual task conditions and attentional load had a much more deteriorating effect on the enumeration of an item than on the detection of an item.

3.3.4   Discussion

This experiment attempted to establish whether dual task situations and attentional load interfere with the detection of a single target and thus, whether the effects on enumeration as seen in Exp. 2.1 arose on the level of target detection or on the level of numerosity retrieval.

The results showed that dual task situations and attentional load also compromised the accuracy of single target detection as apparent in a significant main effect of experimental condition. This replicates and confirms the results found by Joseph, Chun and Nakayama (1997) using a slightly different paradigm and a different feature detection (contrast and orientation odd-ball instead of pure orientation odd-ball). Furthermore, this experiment is an extension to Joseph, Chun and Nakayama (1997) in that it shows that altering the attentional load of the primary task has an affect on single target detection. Thus, even a traditionally pre-attentive defined pop-out visual search task is affected by manipulations that reduce the allocation of attentional resources to the task.

However, the effect in target detection was much weaker than the effect in enumeration, and importantly, in the target present trials, the effect of attentional load was too weak to result in a significant difference between low and high load. As the detection of target absence likely requires differential attentional processing than detection of target presence (Ashbridge, Walsh & Cowey, 1997), the results from target present trials are the crucial ones here when making the comparison to enumeration. The comparison between target present trials of the detection task and numerosity 1 of the enumeration task showed that attentional load had a much more deteriorating effect on enumeration than on detection. Thus, these results suggest that it was mainly the process of numerosity retrieval that was affected by attentional load rather than the process of target detection.

Although the stimulus situations in this comparison were the same (presence of a single high contrast patch), it has to be kept in mind that the task requirements and response possibilities differed. Whereas the target detection task only required the decision between target presence or absence, the enumeration task required the discrimination from seven other numerosities. Nevertheless, it is unlikely that the large differences in effect sizes are entirely due to the difference in task demands. Thus, it appears that the retrieval of numerosity is certainly one, and possibly the main, attention-demanding component in enumeration. The results further suggest that numerosity information is not automatically and concurrently available with item detection as a pre-attentive subitizing mechanism would suggest. The results from both Experiment 2.1 and 2.2 show that it takes processing effort to extract this information from a visual percept and it is therefore unlikely to occur when numerosity is task-irrelevant. In fact, Railo et al. (2007) showed in an inattentional blindness paradigm that enumeration does not occur automatically if it is task-irrelevant.

3.4   General Conclusions

The results of these experiments demonstrated that numerosity judgement, particularly subitizing, is severely impaired when attention is diverted to another attention-demanding task. It is furthermore affected by the attentional load of the additional task, thus specifically by the amount of processing resources available. Although the employed experimental manipulations

also interfere with single target detection, numerosity retrieval from a visual display seems to be the most attention-demanding operation in an enumeration task. Thus, these results provide evidence against a pre-attentive subitizing mechanism and seriously challenge the claim of a privileged attention-independent processing mechanism for small numerosities as proposed by many scientists (e.g. Piazza et al., 2003; Sathian et al., 1999; Simon & Vaishnavi, 1996; Vuilleumier & Rafal, 2000 and others as reviewed in Chapter 1). The fact that dual task conditions and attentional load affected both the subitizing and the estimation range can be taken as an argument that both small and large numerosities may be subserved by a single, continuous enumeration mechanism.

A general criticism could be made with respect to how specific the observed effect of dual task and attentional load is to numerosity. It is conceivable that *any* task is affected by these manipulations, independent of what kind of processing it entails. In fact, this point explains the impact of the study by Joseph, Chun and Nakayama (1997) - if simple feature detection is already attention-demanding, does this mean that every aspect of our perception is attentive? It appears that there are some privileged processes that can be carried out without or with very little attentional resources. For example, Li et al. (2002) have demonstrated that natural scene categorisation (deciding whether a briefly flashed natural scene contains an animal or not) survives dual task situations very well, particularly compared to a simple colour pattern categorisation. In fact, it might be the "gist" of a scene that is perceived also when attention is drawn elsewhere (Wolfe, 1998). The results of Li et al. (2002) are particularly interesting with respect to the fact that natural scene categorisation is thought to involve higher level processing. Thus, whether a judgement can be made with or without attentional resources is not constrained to a certain level of processing, neither with respect to the related brain processing hierarchy nor with respect to the hierarchy in visual processing.

Therefore, although some judgements about properties of a visual scene can be done without or with very few attentional resources, the results from the present experiment suggest that discrete numerosity is not such a property.

Chapter 4

Experiment 3: Neural Correlates of Subitizing under Attentional Load: a fMRI Study

4.1 Introduction

The aim of this experiment was to investigate how attentional load interacted with visual enumeration on a neural level. For this purpose, the dual task paradigm from Experiment 2 was adapted for an fMRI study.

As mentioned in the General Introduction (Chapter 1), most previous studies did not find specific neural activations for subitizing compared to counting. Sathian et al. (1999) presented a grid of horizontal bars (for an unlimited amount of time) in which subjects were asked to either detect the presence or absence of a single vertical bar, or to enumerate the number of vertical bars (1-8). Compared to single target detection, subitizing PET activations were only significantly higher in visual cortex (occipital gyrus). When comparing counting with subitizing, a much more widespread activation pattern was found, including visual and frontal cortex, anterior cingulate and right inferior parietal cortex. Sathian et al. (1999) interpreted these

149

findings as consistent with the idea of pre-attentive, low-level visual processing for subitizing and spatial attentional shifts for counting.

Piazza et al.'s fMRI study mirrored these results (Piazza et al., 2003). They contrasted activations for enumeration of dots (1-7) with colour naming and did not find any greater activations for subitizing compared to colour naming or counting. Counting, however, activated a widespread network in occipital, parietal, insular and prefrontal areas when contrasted with subitizing or colour naming. When analysing the activations in posterior parietal cortex (those that are traditionally associated with numerical processing, c.f. Dehaene et al. (2003), for a review) on a single trial basis, the amount of activation for each numerosity level mirrored the typical behavioural data: little change for numerosities 1 to 3, and increasing activation for numerosities 4 to 7. The authors took these results as evidence for a two-process model and argued for pre-attentive subitizing and serial counting. Also Piazza's earlier PET study (Piazza et al., 2002) did not reveal specific activations for subitizing when contrasted with counting or baseline (enumeration of a single dot), though here the results were interpreted as a continuously increasing activation pattern between small and large numerosities.

The reason why these studies did not find specific activations for subitizing might lie in the applied comparisons. Under the employed experimental conditions (unlimited stimulus presentation times), subitizing is an easy and relatively effortless task. When contrasting brain activations for subitizing with those of a baseline condition comprising an equally effortless task, such as colour naming or single target detection, it may not be surprising that these comparisons did not yield any specific activations. Counting, however (particularly verbal counting), is a much more effortfull and resource demanding task also with unlimited stimulus exposure, thus the emergence of higher level activations when contrasted with effortless task situations would be quite expected. In fact, the only study that discovered specific activations for small numerosities was a recent one by Ansari et al. (2007) which compared small numerosities (1 to 4) with much larger numerosities (10 to 40) under conditions of short stimulus presentation times. Here, the difference in activation levels was sufficiently large to reveal subitizing specific activations in right temporo-parietal junction.

Furthermore, it could be argued that the applied baseline conditions in those earlier

studies contained implicit subitizing as the stimulus was composed of one object (particularly in the baseline conditions applied by Sathian et al., 1999 and Piazza et al., 2002). Thus, it appears that the baseline conditions in these studies were inadequate to find specific activations for subitizing and the comparison with counting a rather unbalanced approach.

4.1.1   Rationale

The current dual task design is very well suited to circumvent the problem of suboptimal baseline conditions. Load theory asserts that under low perceptual load, neural resources can be allocated to distractor processing, i.e. the secondary task. Under high perceptual load, however, distractor processing is suppressed as most neural resources have to be dedicated to the primary task (Lavie, 2005). In the case of the current paradigm, this implies that under single task or low load conditions, sufficient neural resources are available to process numerosity, whereas under high load, neural resources are entirely dedicated to the primary task and numerosity processing is suppressed. Within this logic, colour naming or single target detection (as employed in previous studies) as well as subitizing would both pass as low load tasks, and may thus lead to similar levels of activations as indeed confirmed by previous studies. Under high load, however, numerosity processing is suppressed (as indeed suggested by the behavioural results of Exp. 2.1) and thus provides a much more adequate baseline condition. It was therefore hypothesised that the comparison of neural activations for relatively intact numerosity processing under single task or low load compared to impaired numerosity processing under high load will yield more specific activations for subitizing than previously found.

Furthermore, this design allowed to dissociate the brain regions involved in attentional effort (using the contrasts High Load > Low Load or High Load > Single Task) from those numerosity processing areas that are modulated by attentional load (with the contrasts Low Load > High Load or Single Task > High Load). Thus, this design controls more appropriately for attentional effects of processing demands than previous designs.

By employing numerosities both in the subitizing as well as in the estimation range, the effect of attentional load could be studied in both ranges separately and thus potential differences in activations between small and large numerosities identified, both quantitatively

(i.e. in terms of activation levels) as well as qualitatively (i.e. in terms of brain areas). The numerosity range from Experiment 2 was reduced to the odd numerosities, i.e. numerosities 1, 3, 5 and 7 were employed (in order to have sufficient statistical power on each numerosity level within the amount of trials feasible in a fMRI study). Thus, two numerosity levels for each subitizing and estimation range were available (1 & 3 and 5 & 7, respectively). Three levels of attentional load were applied, no load (single numerosity task), low load and high load (both dual task). The primary colour detection task was not included as a single task, due to time restrictions of the experiment in the MRI scanner.

This design offers many different ways of analysing the data. Only the effects of primary interest are reported here, the applied comparisons and their corresponding predictions were as follows:

1. The effect of numerosity processing (Contrasts: Low Load > High Load or Single Task > High Load). Here, relatively intact numerosity processing across the whole range of numerosities under single task or low load conditions was compared to impaired numerosity processing under high load conditions. Activations were predicted to occur in parietal, occipital and frontal cortex, in line with findings for counting (Piazza et al., 2003) or discrete quantity processing (Castelli et al., 2006). Attentional load modulates neural activity quite early in the visual hierarchy (O'Connor et al., 2002), thus differential activations in occipital cortex for the different load conditions were deemed possible.

2. The effect of subitizing (Contrasts: Subitizing (Low Load) > Subitizing (High Load) or Subitizing (Single Task) > Subitizing (High Load)). Here, only numerosities 1 and 3 are considered and intact subitizing is compared to impaired subitizing. As the results of Experiment 2.1 showed that subitizing is effortfull and attentive, activations in higher level areas outside occipital cortex were predicted, particularly in parietal cortex. Ansari et al. (2007) identified specific activations in the temporo-parietal junction for small numerosities (1-4) compared to large numerosities (10-40), thus this area is a possible candidate for a subitizing specific activation also in this paradigm.

3. The effect of estimation (Contrasts: Estimation (Low Load) > Estimation (High Load) or Estimation (Single Task) > Estimation (High Load)). The same contrast as for subitizing

was carried out selectively for numerosities 5 and 7. Similar activation pattern as previously found for counting and approximate estimation were expected (Piazza et al., 2003 and Piazza et al., 2004). The difference in activation between the subitizing and the counting range could be either quantitatively (i.e. a difference in the strength of activation in the same brain areas) or qualitatively (i.e. different brain areas are activated in each range).

4. The effect of small numerosities (Contrast: Subitizing > Estimation). Small numerosity processing is compared to large numerosity processing across all load conditions, thus independent of load manipulation. This contrast was supposed to identify those areas specific to small numerosities that are not involved in large numerosity processing, as was hypothesised in previous studies. Based on Piazza et al. (2003), no specific activations were predicted.

5. The effect of large numerosities (Contrast: Estimation > Subitizing). Activations specific for large numerosities excluding those for small numerosities were identified across all load conditions. Activations similar to those for counting as found in Piazza et al. (2003) may occur, i.e. frontal and parietal areas. However, as verbal counting was deliberately suppressed in the current paradigm, estimation may involve similar processing than subitizing and thus the difference in activations might not be very pronounced.

6. The effect of attentional effort (Contrasts: High Load > Low Load or High Load > Single Task). These contrasts identified the effect of increased processing demands due to attentional load. Activations of the attentional control network similar to those found in previous studies employing load manipulations were expected, i.e. parietal and frontal areas (Schwartz et al., 2005; O'Connor et al. 2002).

7. The effect of dual task (Contrast: Low Load > Single Task). This contrast was supposed to identify the effect of increased processing demands due to dual task situation. Areas specific for the primary colour detection task were expected to be found, possibly areas involved in colour detection and object recognition (occipital and temporal cortex).

Note that both for the effects of numerosity processing as well as for the effects of attentional effort, parietal and frontal areas were predicted. It lies in the nature of this design

that activations reflecting *both* numerosity processing (as identified by employing the load manipulation) *and* attentional effort could not be identified. For example, the contrasts High Load > Low Load (attentional effort) and Low Load > High Load (numerosity processing) yielded, by definition, mutually exclusive activations. At the same time, however, it was a strength of the design to be able to separate the activations related purely to numerosity processing from those of attentional effort.

With respect to the original hypothesis of this thesis, it was predicted that if subitizing is pre-attentive and recruits only low level processing, the contrasts Subitizing (Low Load) > Subitizing (High Load) or Subitizing (Single Task) > Subitizing (High Load) should either not yield any specific activations, or only activations in visual areas. If subitizing is attentive, however, the same contrasts were predicted to result in activations of higher level areas such as parietal and frontal cortex (Corbetta & Shulman, 2002).

The same visual stimulus as in Experiment 2 was used, i.e. a central colour target for the primary task and circularly arranged Gabor-like patches for the secondary numerical task. However, as parietal cortex responds to peripheral visual stimulation and this response is modulated by attentional load (Schwartz et al., 2005), the size of the stimulus was decreased in order to prevent the confound of increasing parietal signal due to increasing processing demands in the periphery with target number increase. It was hypothesised that with a smaller visual stimulus, parietal cortex activation should be more specific to numerosity processing and not purely related to peripheral visual stimulation.

4.2   Methods

4.2.1   Subjects

18 subjects (mean age: 24.5, range: 19 - 35, 10 females) participated in the experiment. Subjects were screened for right-handedness, intact colour vision and MRI suitability, i.e. absence of neurological or psychiatric illnesses and absence of metal in the body. All signed informed consent and were paid for their participation. The study was approved by the ethics committee

of the Dept. of Psychology at UCL.

4.2.2   Stimuli

The visual stimuli were the same as in Experiment 2, though smaller in size. The central diamond shape subtended 1.6°, the circle of Gabor-like patches 6.5° and each patch 1°. Stimuli were generated using the Cogent toolbox (www.vislab.ucl.ac.uk/Cogent/) for MATLAB (Mathworks, Inc).

4.2.3   Task and Experimental Procedure

The primary colour detection task was the same as in Experiment 2. The secondary numerical task required subjects to judge whether there were 1, 3, 5 or 7 high contrast Gabor-like patches in the circle. The total number of items in the circle ranged from 9 to 12, counterbalanced for each target number and load condition. Three experimental conditions were employed: single numerosity task (no load), dual task (low load) and dual task (high load). The fMRI design was blocked with respect to the experimental condition and event-related with respect to numerosity levels.

Each experimental block started with a screen displaying the instruction for the load condition ("Numerosity only" for single task, "Red" for low load, " Green right-tilted or yellow left-tilted" for high load), the first trial of the block was then initiated after 6 repetition times (21.6 s). As in Experiment 2, a fixation cross was displayed for 1 second, followed by the stimulus for 200 ms and subsequently followed by a mask. Subjects were given a maximum of 2.5 seconds to respond to the primary task and a maximum of 3.5 seconds to respond to the secondary task (due to time restrictions in the scanner). If subjects did not respond within this time, the next trial was started and a missing value recorded. Inter-trial intervals were 2 repetition times (7.6 s) within a block.

A right-hand key pad with 4 response buttons was used. Under dual task conditions, subjects responded always first to the primary task with index and middle finger (indicating target presence or absence) and subsequently to the secondary numerical task using all four

buttons (index finger to little finger) indicating target numbers 1, 3, 5 and 7. Under single task conditions, only one response to the numerical task was required. Speed was emphasised for the primary task and accuracy for the secondary task. Subjects were prompted to take breaks after each run. The scanning session lasted for about 1 hour. Prior to the scanning session, each subject received training with 2 blocks (16 trials each) of each experimental condition (96 trials in total).

4.2.4   Experimental Design

4 runs of 6 blocks (2 for each experimental condition) were employed. Each block consisted of 16 trials, thus the whole experiment comprised 384 trials in total, 32 trials for each of the 12 numerosity x load combinations (4 numerosities, 3 experimental conditions). The order of experimental condition was counterbalanced within each run. Numerosity levels were counterbalanced across the 2 blocks of each experimental condition in each run and the order of numerosity levels randomised.

4.2.5   Pilot Experiment

Prior to scanning, a pilot experiment was carried out in order to establish whether the manipulation of dual task and attentional load also worked with the reduced range of numerosities and with the smaller stimulus size. 7 subjects were tested under the same experimental conditions as in Experiment 2.1. Mean accuracy data mirrored the effect found in Experiment 2: Subjects performed well in the single task condition, but were increasingly impaired under low load and again under high load conditions (Main effect: $F(2,12) = 38.01$, $p < .001$; all post-hoc comparisons $p < .05$). There was an effect of target number ($F(3,18) = 19.39$, $p < .001$), but no interaction ($F(6,36) = .22$, $p > .05$). Thus, also under the altered task and stimulus conditions, the experimental manipulation of dual task and load affected enumeration as in Experiment 2.1.

4.2.6   Image Acquisition

Subjects were scanned at the Birkbeck-UCL Neuroimaging Centre in a Siemens TIM Avanto 1.5 Tesla MRI scanner. Functional images measuring BOLD signal were acquired with a gradient-echo EPI sequence (38 slices, TR = 3.6 sec; TE = 50 ms, FOV = 192 x 192, matrix = 64 x 64) giving a notion resolution of 3 x 3 x 3 mm. Between 230 and 300 volumes for each run were recorded (depending on how quickly individual subjects responded to the task). Reduced signal was recorded in the frontal lobe due to failure of the frontal head coil in about half of the scanning sessions.

4.2.7   Data Analysis

The behavioural data were analysed as in Experiment 2. The imaging data were analysed using statistical parametric mapping (SPM 5; www.fil.ion.ucl.ac.uk/spm/). Functional images were submitted to a random-effects group analysis using a general linear model applied at each voxel across the whole brain, according to standard SPM processing (Friston et al., 1995). The first five scans of each run were discarded to allow for T1 calibration effects. At pre-processing, all scans for each subjects were spatially realigned to the first image of the first run, time-corrected with respect to the middle slice and normalised to a standard anatomical template (resampled voxel size: 2 x 2 x 2 mm^3). Scans were smoothed with an isotropic 8 mm full-width half-maximum (FWHM) Gaussian kernel. Time-series from each voxel were high-pass filtered (1.120 Hz cutoff) to remove low-frequency noise. Random-effects statistical analysis was carried out as two stages of a mixed-effects model.

In the first subject-level analysis, the 12 conditions of interest (3 load conditions x 4 numerosity levels) were modeled by stick functions convolved with a canonical hemodynamic response function (HRF) and used as covariates in a multiple regression analysis. Parameter estimates for each covariate were estimated at each voxel by a least-squares fit to the data of each condition and each individual subject. Statistical parametric maps of the t-statistic were generated from linear contrasts testing the main effects in each subject. These individual parameter estimates were then included in a second level group analysis using a 3 x 4 ANOVA.

This resulted in a random effect SPM t statistic for each comparison of interest, thresholded voxel-wise at $p = .05$ with family-wise error (FWE) correction for multiple comparison. At some regions of interest, small volume correction was applied by selecting a spherical space of 10 mm around the region of interest and adjusting p values to this volume.

4.3 Results

4.3.1 Behavioural Results

Colour Detection Task

Mean accuracy and reaction time data for the primary colour detection task are depicted in Fig. 4.1 for both load conditions. Subjects were significantly less accurate under high load than under low load (t = 5.82, $p < .001$) and significantly slower (t = -14.76, $p < .001$). Thus, as in Experiment 2, the manipulation of attentional load was effective.

Numerosity Judgement Task

Accuracy As can be seen in Fig. 4.2, accuracy data mirrored the results from Experiment 2. Subjects were significantly less accurate under low load conditions compared to single task conditions, and again less accurate under high load than under low load conditions (Main effect: F(2,34) = 78.60, $p < .001$, all post-hoc comparisons $p < .001$). As expected, accuracy decreased with target number (F(3,51) = 60.91, $p < .001$), but no interaction between condition and target number was observed (F(6,102) = .89, $p > .05$), contrary to Experiment 2.

Mean Responses Also in mean response data (Fig. 4.3, top panel), there was again a significant main effect of condition (F(2,34) = 12.84, $p < .001$), although the difference between low load and high load conditions was not significant in the post-hoc comparison ($p = .077$). Both low load and high load condition, however, differed from the single task condition ($p = .040$ and $p = .001$, respectively). Underestimation occurred for numerosities 5 and 7 in all conditions (t $\leq$
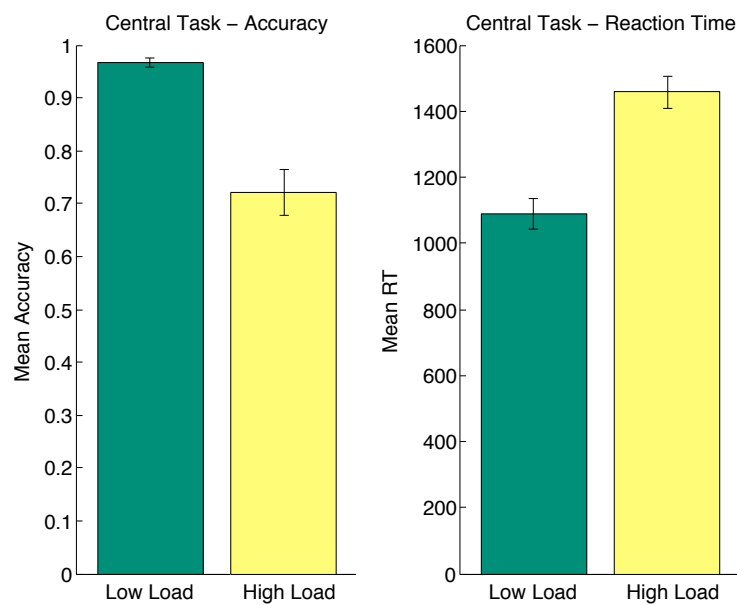
Figure 4.1: Exp. 3 Behavioural Results - Central Colour Task.

Behavioural results of Exp. 3 for the central colour detection task. Mean accuracy (left) and mean reaction times (right) are plotted for both low and high load condition. Error bars indicate 1 SEM.
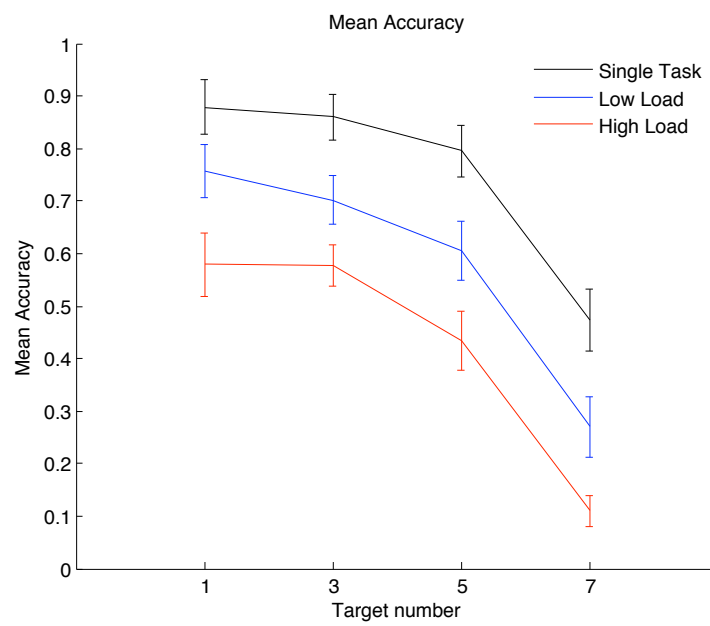
Figure 4.2: Exp. 3 Behavioural Results - Numerosity Task - Accuracy.

Accuracy data (proportion correct) of the numerosity task of Exp. 3 as a function of target number for each load condition. Error bars indicate 1 SEM.

-3.88, $p < .001$), whereas numerosity 3 exhibited neither under- nor overestimation (t $\leq$ 1.82, $p > .05$). At numerosity 1, overestimation occurred for both load conditions (t $\geq$ 4.49, $p < .001$), but only marginally in the single task condition (t $=$ 2.02, $p = .059$). As expected, there was a significant effect of target number (F(3,51) $=$ 96.25, $p < .001$) that interacted with condition (F(6,102) $=$ 47.41, $p < .001$).

Response Variance   Mean response variance (Fig. 4.3, middle panel) increased with increasing attentional load (main effect: F(2,34) $=$ 19.53, $p < .001$), though the difference between low and high attentional load was not significant in the post-hoc comparison ($p > .05$). However, response variance was significantly lower in the single task condition compared to both load conditions ($p < .001$). Variances increased with target number (F(3,51) $=$ 8.76, $p < .001$), but this effect did not interact with condition (F(6,102) $=$ .81, $p > .05$).

Weber Fraction   Weber fractions (Fig. 4.3, bottom panel) gave a similar picture than in Experiment 2.1. Weber fractions were highest in the high load condition and decreased in low load and single task conditions (main effect: F(2,34) $=$ 28.73, $p < .001$; post-hoc comparisons: single task versus low load: $p < .001$, single task versus high load: $p < .001$, low versus high load: $p = .016$). There was a main effect of target number (F(3,51) $=$ 54.34, $p < .001$) that interacted with condition (F(6,102) $=$ 17.72, $p < .001$). When conditions were considered separately, Weber fractions in the single task condition changed with target number (F(3,51) $=$ 8.03, $p < .001$), contrary to Experiment 2.

4.3.2   Imaging Results

Effect of Numerosity Processing

Here, activations for low load and single task were compared to high load activations across the whole numerosity range. The contrast Low Load > High Load did not yield any significant activations when p values were adjusted for multiple comparisons. The contrast Single Task > High Load yielded significant activations in right middle frontal gyrus, right intraparietal lobule,
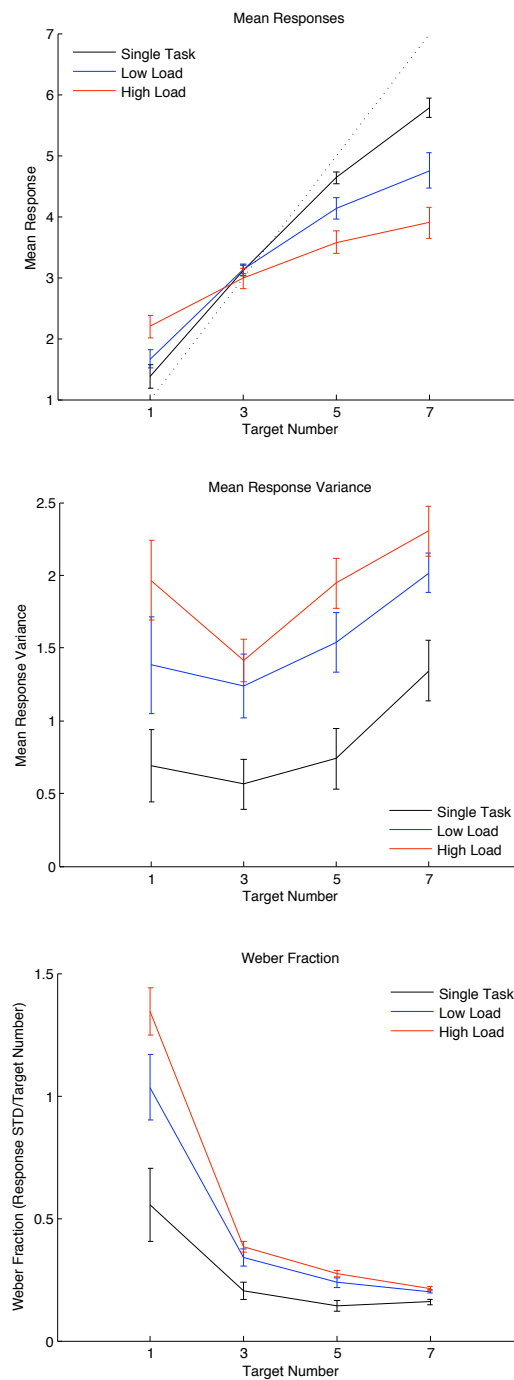
Figure 4.3: Exp. 3 Behavioural Results - Numerosity Task

Mean responses (top), mean response variances (middle) and Weber fraction (bot-
tom) of Exp. 3 as a function of target number. The diagonal line in the graph for
mean responses indicates perfect performance. Error bars reflect 1 SEM.

Imaging Results - Effect of Numerosity Processing (Single Task > High Load)

Side	Area	X	Y	Z	Z value	p value	Sub	Est
R	Middle frontal gyrus	18	32	42	5.63	<.001	*	*
		30	38	46	4.75	.026	*	
		28	40	44	4.68	.034	*	
R	Intraparietal lobule	58	-42	52	5.27	.002	*	*
L	Cingulate gyrus	-14	-18	42	5.04	.007	*	*
R		10	-26	30	5.03	.008	*	*
		8	-24	28	4.73	.028	*	*
		8	-32	34	4.63	.042	*	*
L	Transverse temporal gyrus	-34	36	12	4.93	.012	*	*
R	Thalamus	14	-18	-2	4.92	.012	*	
		18	-16	2	4.59	.050	*	*

Table 4.1: Exp. 3 Imaging Results - Effect of Numerosity Processing

Neural activations for the contrast Single Task > High Load across the whole numerosity range. Coordinates are in MNI space. p values are corrected for multiple comparisons (family-wise error (FWE)). The last two columns indicate significant activation with small volume correction at contrasts Subitizing (Single Task) > Subitizing (High Load) or Estimation (Single Task) > Estimation (High Load), respectively.

bilateral cingulate gyrus, left transverse temporal gyrus as well as right thalamus (summarised in Table 4.1). All these areas were analysed separately with small volume correction in the contrasts Subitizing (Single Task) > Subitizing (High Load) and Estimation (Single Task) > Estimation (High Load) in order to see whether these areas were activated as well in both numerosity ranges separately. All areas showed significant activations in the subitizing contrast, but some activations in right middle frontal gyrus and left transverse temporal gyrus did not reach significance in the estimation contrast (see Table 4.1).

Effect of Subitizing

The contrast Subitizing (Low Load) > Subitizing (High Load) yielded one specific activation in right temporo-parietal junction / superior temporal sulcus at MNI coordinates (40 -54 22) (Z = 4.94, p = .011), as depicted in Fig. 4.4. No other significant activations were found.

Figure 4.5 depicts the signal change at temporo-parietal junction (40 -54 22) averaged across the whole experiment for each of the different numerosity and load combinations. Activation levels were clearly numerosity modulated in the single task and low load conditions, but not in the high load condition. This is consistent with the idea that numerosity processing is suppressed under high load. Overall activation levels between single task and low load were similar, whereas high load yielded lower activation levels.[1]

The contrast Subitizing (Single Task) > Subitizing (High Load) yielded one specific activation in right middle frontal gyrus (20 36 42)(Z = 4.96, p = .010, Fig. 4.6), very close to the area identified when considering the whole numerosity range in the contrast Single Task > High Load (see Table 4.1). Signal change at this area reflected a different pattern at this location than in temporal parietal junction: the signal decreased consistently with increasing numerosity and load condition (Fig. 4.7).

Effect of Estimation

Neither the contrast Estimation (Low Load) > Estimation (High Load) nor the contrast Estimation (Single Task) > Estimation (High Load) yielded any significant activations at the corrected level. When analysing both contrasts with small volume correction for the area in right temporo-parietal junction as identified in the subitizing contrast, again no significant activation was found. For the area in right middle frontal gyrus, however, small volume correction identified a significant activation very close by at (18 32 40)(Z = 3.42, p = .020). This activation belonged to the same cluster as found when analysing the whole numerosity range.

---

[1]Note that the baseline level at 0 is a consequence of the SPM algorithm and rather arbitrary, thus negative values do not necessarily reflect suppression, nor do positive values reflect activation. The relative difference between activation levels is the meaningful measure here.
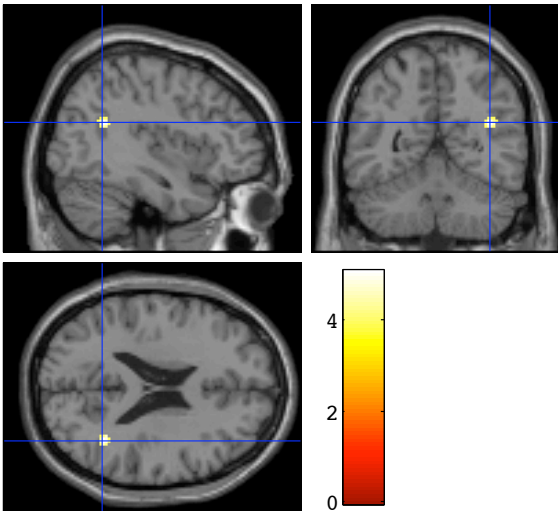
Figure 4.4: Exp. 3 Imaging Results - Effect of Subitizing

Neural activations for the effect of subitizing as revealed in the contrast Subitizing (Low Load) > Subitizing (High Load). The significant activation ($p = .011$, FWE corrected for multiple comparisons) is in right temporo-parietal junction / superior temporal sulcus (40 -54 22). The colour scale indicates Z value.
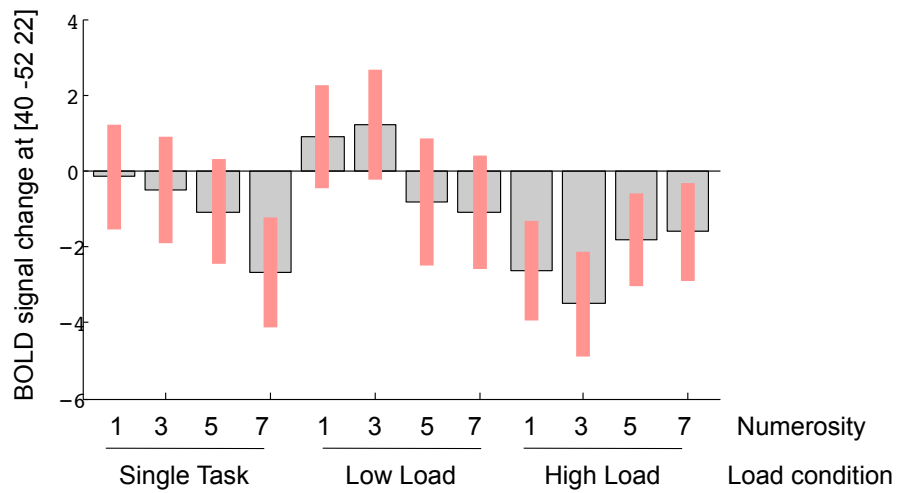
Figure 4.5: Exp. 3 Imaging Results - Signal Change at Right Temporo-parietal Junction

BOLD signal change at right temporo-parietal junction (40 -54 22) for each numerosity and load level. Error bars indicate 90% confidence intervals.
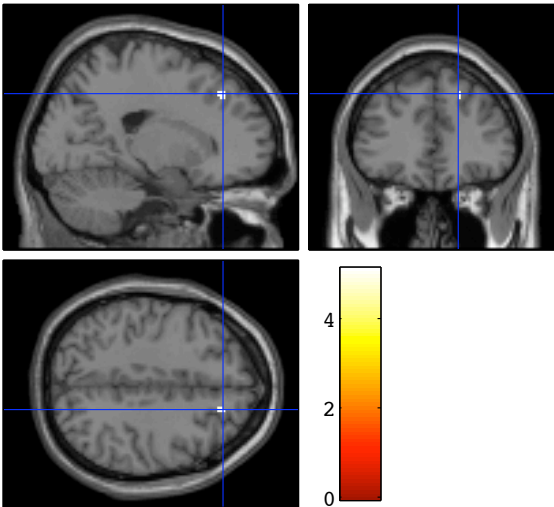
Figure 4.6: Exp. 3 Imaging Results - Effect of Subitizing

Neural activations for the effect of subitizing as revealed in the contrast Subitizing (Single Task) > Subitizing (High Load). The significant activation ($p = .010$, FWE corrected for multiple comparisons) is in right middle frontal gyrus (20 36 42). The colour scale indicates Z value.
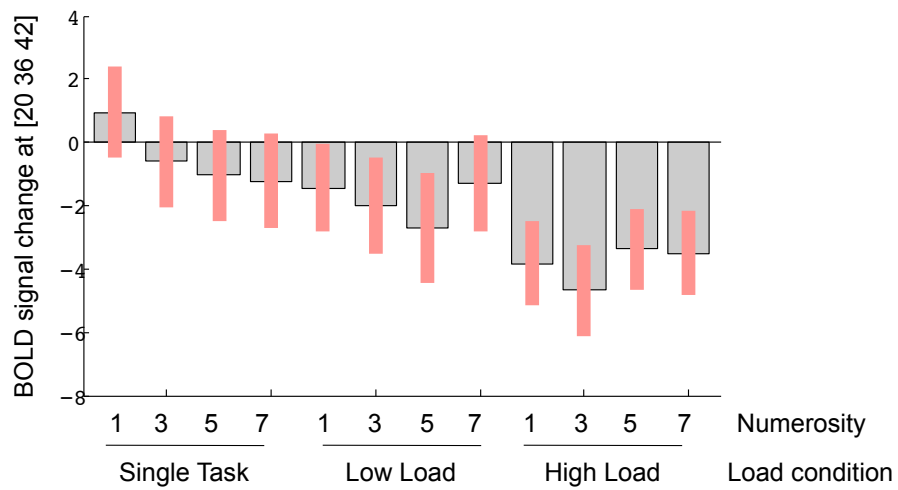
Figure 4.7: Exp. 3 Imaging Results - Signal Change at Right Middle Frontal Gyrus

BOLD signal change at right middle frontal gyrus (20 36 42) for each numerosity and load level. Error bars indicate 90% confidence intervals.

Imaging Results - Effect of Attentional Effort (High Load > Single Task)

Side	Area	X	Y	Z	Z value	p value
L	Precuneus	-20	-64	20	6.31	<.001
L	Lingual gyrus	-16	-44	0	5.73	<.001
L	Parahippocampal gyrus	-22	2	-12	5.47	.001
R		26	-20	-10	5.39	.001
		22	2	-10	4.91	.013
R	Posterior cingulate	26	-58	20	4.71	.030

Table 4.2: Exp. 3 Imaging Results - Effect of Attentional Effort

Neural activations for the contrast High Load > Single Task. Coordinates are in MNI space. p values are corrected for multiple comparisons (family-wise error (FWE)).

Effect of Small and Large Numerosities

Neither the contrast Subitizing > Estimation nor the contrast Estimation > Subitizing yielded any significant activations at the corrected level. The same was true when the same contrasts were applied to each load condition separately.

Effect of Attentional Effort

The contrast High Load > Low Load did not yield any significant activations at corrected level. When comparing high load activations with single task activations (contrast High Load > Single Task), significant responses in left precuneus, left lingual gyrus, bilateral parahippocampal gyrus as well as right posterior cingulate cortex were found (see Table 4.2).

Effect of Dual Task

Low load activations were compared with single task activations and yielded significant responses in right superior temporal gyrus (36 18 -28) (Z = 4.77, p = .023) and left parahippocampal gyrus

(-28 -24 -18) (Z = 4.61, $p = .045$).

4.4   Discussion

This experiment investigated the neural basis of the interaction between attentional load and visual enumeration. Based on Lavie's load theory (Lavie, 2005) it was predicted that neural activity related to numerosity processing is suppressed under high attentional load compared to low attentional load and single task conditions. Thus, suppressed activation levels under high attentional load were hypothesised to provide an adequate baseline condition to identify neural activations related to intact numerosity processing, specifically subitizing. Furthermore, it was predicted that subitizing should involve areas associated with higher level, attentional processing if it was truly attentive.

The behavioural results confirmed that the manipulation of attentional load on enumeration could be replicated under imaging conditions and with the slightly changed stimulus display and target number ranges. Again, enumeration accuracy was decreased under dual task conditions when only a single colour feature had to be detected in addition to the enumeration task. With increasing attentional load, enumeration accuracy was even more impaired. There was no interaction between target number and the effect of condition, suggesting that attentional load affected all numerosities to an equal extent. Analysis of mean responses and mean variances revealed that the difference between low load and high load was not as pronounced as in Experiment 2, suggesting that subjects' choice of responses was similarly variable in both load conditions.

4.4.1   The Effects of Subitizing and Estimation

For this thesis, the activations specific to subitizing and their potential difference to estimation were of primary interest and are thus discussed first.

Activation in Temporo-parietal Junction

The behavioural results indicate that, in line with Lavie's load theory, subitizing was suppressed under high attentional load. Indeed, the contrast Subitizing (Low Load) > Subitizing (High Load) revealed one specific activation in the right temporo-parietal junction (TPJ). Importantly, this activation was specific to the subitizing range and did not emerge in the estimation range, even when small volume correction was applied at the identified location. This area therefore seems to be specifically activated when attentional load modulates small numerosity judgement.

Almost exactly the same area was identified by Ansari et al. (2007) when comparing activations related to the enumeration of 1 to 4 dots with the estimation of 10 to 40 dots. Their contrast Small Numerosities > Large Numerosities also revealed a single specific activation at (41, -56, 23), very close to the one identified here for the contrast Subitizing (Low Load) > Subitizing (High Load) at (40, -54, 22). Critically, Ansari et al. (2007) also did not find TPJ activation in the opposite contrast when comparing large numerosities with small numerosities. They even report a suppression in TPJ relative to baseline in the large numerosity range. [2]

The right TPJ has been associated predominantly with stimulus-driven attention and is involved in target detection and reorienting attention towards salient or unexpected sensory events (e.g. Corbetta et al., 2000; Marois, Leung & Gore, 2000; Serences et al., 2005). TPJ activation is particularly pronounced when these events are behaviourally relevant (Kincade et al., 2005). According to Corbetta and Shulman's (2002) influential theory on the neural basis of visual attention, the right TPJ is an important part of a right lateralised ventral frontoparietal network responsible for stimulus-driven, bottom-up attentional control. This network is distinct from a bilateral, more dorsally located network responsible for goal-driven, top-down attention which includes intraparietal and superior frontal areas.

Based on this evidence, it could be argued that TPJ activation arose due to the target stimulus being briefly flashed. As such the stimulus was both salient, slightly unexpected with regard to its onset and behaviourally relevant. However, this was true for the stimulus containing both small as well as large target numbers, and as TPJ activation was selectively found for small

---

[2]It is not clear, however, how the baseline condition was defined in Ansari et al. (2007).

numerosities, it is unlikely to have arisen purely due to sudden onset of the stimulus. Quite to the contrary, given that more behaviourally relevant targets were present in the large numerosity stimuli, TPJ activation would be expected to occur stronger (or at least equally) pronounced in the estimation range. From the present results as well as those by Ansari et al. (2007), it appears that TPJ activation is modulated by the number of items that are behaviourally relevant, and is indeed suppressed at higher numerosities. In the present experiment, the BOLD signal change at TPJ (Fig. 4.5) shows that its activation is clearly numerosity modulated and increasingly suppressed with increasing numerosity in the single task and low load conditions. In the high load condition, TPJ activation is equally suppressed in all numerosities which can be taken as evidence that the load manipulation is powerful enough to suppress activity efficiently also in the subitizing range (as would be predicted by load theory (Lavie, 2005)). The similar extent of suppression in the estimation range at low load as well as high load explains why the contrast Estimation (Low Load) > Estimation (High Load) did not yield a significant response at this particular location. Together with the findings from Ansari et al. (2007), these results show that TPJ activity appears to be specifically modulated by the number of items the brain deals with. TPJ is active during small numerosity judgement and is increasingly suppressed at higher numerosities.

Interestingly, Todd, Fougnie and Marois (2005) found that increasing visual short-term memory load selectively suppresses TPJ activity. They report a slightly more lateralised activation site at (59, -47, 24). Although they employed a rather different paradigm (change detection, as typically used to probe visual short-term memory capacity), TPJ activity was suppressed with increasing number of items to be memorised. Crucially, a modulation of TPJ suppression was found between set size 1 and set size 3, but not between set size 3 and 6. Thus, also here, TPJ modulation was specific to the small number of items to be handled and ceased at higher set sizes. Given that TPJ is also involved in attentional processes, Todd, Fougnie and Marois (2005) proposed that the VSTM load-related suppression is likely to reflect the attentional demands for VSTM maintenance. However, this is not quite consistent with the evidence that TPJ is related to detection of salient sensory events and stimulus-driven attention. Following Corbetta and Shulman's theory, maintaining a memory for a visual attributes is mediated by the dorsal fronto-parietal top-down network (as evidenced in Corbetta, Kincade &

Shulman, 2002, for example) and a such does not involve TPJ. Thus, it might be more plausible to conclude that stimulus-driven TPJ activity is modulated by the number of items the brain has to handle.

Given that TPJ is associated with many different processes related to stimulus-driven attention, it would not be wise to conclude that TPJ activation is purely specific to numerosity processing or subitizing in particular. Based on the evidence from this study, together with the findings from Ansari et al. 2007 and Todd, Fougnie and Marois (2005), TPJ activity appears to depend on the number of behaviourally relevant stimuli that elicit bottom-up attention.

It would also not be wise to conclude that small and large numerosities are subserved by distinct neural substrates, purely on the basis that there is a brain area differentially responsive when the number of target items is small (indeed, a similar assumption has been made on large numerosities in Piazza et al. (2003), which is a questionable approach as discussed in Chapter 1). A similar argument as made in the General Introduction related to Weber's law could be raised, such that the resolution of the visual system reaches its threshold at about 5 items and handling more items requires larger differences in stimulus size. The relative difference between 5 and 7 items might be too small to be differentially treated by the visual system or, alternatively, the relative difference in activation for these stimuli is too small to be extracted by the fMRI method. Indeed, when contrasting small numerosities with large numerosities across load conditions or even within each load condition, no area could be isolated that was preferentially activated by small or large numerosities.

Activation in Middle Frontal Gyrus

The contrast Subitizing (Single Task) > Subitizing (High Load) revealed a specific activation in right middle frontal gyrus (MFG) at (20, 36, 42). This area was also identified when applying the same contrast across the whole numerosity range (Single Task > High Load) and when applying small volume correction at the contrast Estimation (Single Task) > Estimation (High Load). The signal change at MFG showed that also this area is modulated by numerosity in the single task and low load condition, and activation is suppressed in the high load condition, as at the TPJ site. In contrast to TPJ, however, activation levels at MFG decrease fairly linearly from

173

single task to low load to high load conditions and are not on comparable levels in single task and low load as in TPJ. This linear decrease in suppression explains why this area in MFG only revealed significant activations in the contrast Single Task > High Load and not in the contrast Subitizing (Low Load) > Subitizing (High Load). Fig. 4.7 shows that the relative differences between low and high load conditions are not pronounced enough to yield a significant difference. Similarly, Fig. 4.5 shows that, on average, activation levels in rTPJ are similar between single task and high load conditions. Thus, no significant rTPJ activations were found in the contrast Single Task > High Load.

Right middle frontal gyrus is consistently recruited during detection of novel or low-frequency events and also part of the ventral fronto-parietal network responsible for stimulus-driven attention (Corbetta & Shulman, 2002). As such, this activation is very likely related to the effect seen in temporo-parietal junction. In contrast to the TPJ site, however, the modulation of MFG is not restricted to the small numerosities but occurs over the whole numerosity range. In fact, this area is activated in the contrast applied to the whole numerosity range and also, to a weaker extent, in the contrast for the estimation range. Thus, MFG seems to be related more generally to numerosity processing, and given the linear suppression from single task to high load, possibly to task difficulty.

Taken together, with regard to the hypothesis of this thesis, the specific activations in temporo-parietal junction as well as middle frontal gyrus are evidence for the involvement of the ventral fronto-parietal attention network in small numerosity judgement, and as such another, neurally grounded argument against a pre-attentive subitizing mechanism. Thus, these results link small numerosity processing with stimulus driven attention mechanism on a neural level.

The results furthermore explain why many previous studies did not find specific activation for subitizing. As predicted, the high load manipulation was sufficiently effective to suppress numerosity related processing in all numerosity ranges and thus provided an appropriate baseline condition. Under single task and low low load conditions, however, BOLD signal change levels for subitizing were relatively low, leading to a significant effect only when compared to the strongly suppressed subitizing levels under high load. Previous studies presumably compared the low activation levels of subitizing under single task with similarly low activation

levels of single target detection or colour naming, and thus failed to find specific activations for subitizing.

The Comparison of Small versus Large Numerosities

The graphs in Fig. 4.5 and Fig. 4.7 also demonstrate why the two numerosity modulated areas TPJ and MFG did not yield significant results when subitizing and estimation ranges were contrasted between each other (contrasts Subitizing > Estimation and Estimation > Subitizing). Although in both areas signal change decreases with numerosity, the relative differences in activation levels between the small and large numerosities are not pronounced enough to yield significant differences between both ranges. In fact, the modulation suggests that there might be a rather smooth transition between small and large numerosities without a clear-cut discontinuity. However, this was not formally tested. The fact that no other area responded selectively when contrasting small with large numerosities indicates that the difference in neural response between both ranges cannot be very pronounced, at least as measured in the current paradigm. This results supports the idea that there is no functional segregation for the enumeration of small and large numerosities.

The reason why previous studies found differences for these contrasts might have lied in the applied paradigms. Many of the older studies have used long stimulus presentation times, thus facilitating subitizing and allowing verbal counting in the higher numerosities (Piazza et al., 2002; Piazza et al., 2003; Sathian, 1999). It is possible that the activations found for Counting > Subitizing in these studies were more related to verbal counting than to estimation. In the paradigm employed here, short stimulus presentations ensured that subjects also estimated the higher numerosities and the behavioural data suggest that the load manipulation affected all numerosities to a similar extent. As such, the difference in processing might not have been pronounced enough to yield differential activations. The reason why Ansari at al. (2007) was the only study to find significant activations for the contrast Small Numerosities > Large Numerosities (in right TPJ) might have been due to the fact that they contrasted subitizing with the estimation of much larger collections (10 to 40 dots), thus achieving much higher differences in processing demands than by contrasting the subitizing range with the immediately following

numerosities from 5 onwards.

4.4.2   The Effect of Numerosity Processing across the Whole Range

The contrast (Single Task > High Load) when applied across the whole numerosity range yielded a network of activations mainly in right middle frontal gyrus, right intraparietal lobule and cingulate gyrus. The strongest activation in MFG was the same as found in the subitizing and, with small volume correction, in the estimation contrast, reflecting numerosity processing or possibly task difficulty as discussed above. Again, the emergence of several activations in middle frontal gyrus suggests that the numerosity task involved the ventral network of stimulus-driven attention (Corbetta & Shulman, 2002).

The activation in the right intraparietal lobule is likely to be related to numerosity processing, too. Ansari et al. (2007) found activations in inferior parietal lobule for estimating large numerosities when contrasted with small numerosities. Intraparietal responses have been found in most previous enumeration studies for counting (Piazza et al., 2002; Piazza et al., 2003; Sathian et al., 1999), for both discrete and approximate quantity processing (Castelli et al., 2006) as well as for approximate numerosity estimation (Piazza et al., 2004).

Cingulate cortex has been shown to be involved in decision making in general (e.g. Forstmann et al., 2008) and specifically in the activation of correct responses and the inhibition of incorrect responses in a numerical stroop task (Bush et al., 1998). Thus, the activation in this area could be related to the selection of numerical responses.

The fact that the contrast (Low Load > High Load) did not yield any significant activations on the corrected level suggests that numerosity processing was already relatively suppressed under low load and the difference to high load not pronounced enough. In fact, the results from the behavioural data with respect to response behaviour (i.e. mean response and response variance) demonstrated that subjects responded similarly variably in both load conditions.

4.4.3   The Effect of Dual Task and Attentional Effort

The opposite contrasts than those applied for numerosity processing reflect the effects of attentional effort and dual task, i.e. those areas mainly involved with the increasing attentional demands of performing a concurrent colour detection task as well as the attentional demands of increasing load. These effects are supposed to be independent of numerical processing. Unexpectedly, the contrast High Load > Low Load did not yield significant activations, contrary to other studies manipulating attentional load (e.g. Schwartz et al., 2005). However, the contrast High Load > Single Task yielded activations in attentional areas (precuneus and lingual gyrus) and in parahippocampal gyrus, classically associated with object recognition (e.g. Maguire et al., 1998) and colour detection (e.g. Dojat et al., 2006). Given that the high load condition requested subjects to detect a certain colour-orientation combination (involving retrieval from memory), this was a predicted effect. Also the contrast Low Load > Single Task activated parahippocampal gyrus, suggesting that detecting the red coloured triangles in the low load conditions also recruited object recognition processes. Furthermore, this contrast identified activation in superior temporal gyrus which is one of the key sites of damage in hemi-spatial neglect (Karnath, Ferber & Himmelbach, 2001) and, according to Corbetta and Shulman (2002), part of the ventral stimulus-driven attentional network. This indicates that performing two tasks instead of one probes attentional processing, as would be predicted from load theory (Lavie, 2005).

4.4.4   Conclusions

Due to the manipulation of attentional load in the current dual task paradigm, specific activations for subitizing were identified in brain areas classically associated with stimulus-driven attention. Thus, these findings link subitizing with attentional processes and provide further evidence against a pre-attentive subitizing mechanism. The absence of activations specific to small or large numerosities suggests that, under controlled task conditions, subitizing and estimation may not necessarily be subserved by two functionally distinct neural mechanisms as proposed in previous studies (e.g. Piazza et al., 2003)

Chapter 5

General Discussion

5.1 Summary of Results

In this thesis, the role of attention in visual enumeration was studied, with particular emphasis on the subitizing process, i.e. the judgement of small numerosities. Specifically, the hypothesis of a parallel and pre-attentive subitizing mechanism was tested as traditionally proposed in many enumeration studies (e.g. Sagi & Julesz, 1985; Dehaene & Cohen, 1994; Simon & Vaishnavi, 1996; Sathian et al., 1999; Vuilleumier & Rafal, 2000; Piazza et al., 2003).

In this work, it was not attempted to infer the attentional demands of the subitizing process from performance slopes as done by many previous studies (e.g. Trick & Pylyshyn, 1993) since this has been shown to be an insufficient measure of the attentional nature of a task (c.f. DiLollo et al., 2001). Instead, the amount of attention that could be allocated to the enumeration task was experimentally manipulated.

The results from all experiments fail to support the hypothesis of a parallel and pre-attentive subitizing mechanism and instead suggest that subitizing is an attentive process. Experiment 1 demonstrated that subjects are unable to correctly retrieve the numerosity from two spatially intermixed subsets in parallel. Subjects have a clear advantage when they know which subset to attend to (the case of pre-cueing) and a clear disadvantage if attention needs to be

distributed to two sets (the case of post-cueing). Attentional set as manipulated by pre- and post-cueing affects both subitizing and estimation to a similar extent. Experiment 2 showed that imposing an additional, attention-demanding task to an enumeration task impairs enumeration in both the small and large numerosity ranges. Furthermore, the attentional requirements, i.e. the load of this additional task selectively impairs enumeration. That is, the higher the attentional load of the additional task, the greater the effect on enumeration. Thus, the more attentional resources are taken away from an enumeration task, the more numerosity judgement, particularly subitizing, deteriorates. Experiment 3 showed that relatively intact subitizing compared to impaired subitizing selectively activates brain areas associated with stimulus-driven attention. This suggests that intact subitizing relies on attentional processing.

The results furthermore demonstrate that numerosity judgement depends on other processes such as single target detection, target-distractor segregation, perceptual grouping and possibly feature binding. These processes can potentially be attention-demanding and thus have an impact on the attentional requirements of enumeration. Therefore, it appears that small numerosity information is not an automatically extracted feature from the visual environment.

Given that the employed attentional manipulations affected both subitizing and estimation to a similar extent, the current results indicate that the idea of an attentional dichotomy between small and large numerosity judgement is flawed. As the attentional dichotomy is one of the main arguments for a functional dichotomy, also the existence of two functionally distinct enumeration mechanisms is put into question. Thus, the idea of a strict set-size limit at 4 items, and as such the proposal of a hard-wired "core system" for small numerosities (e.g. as suggested by Feigenson, Dehaene & Spelke, 2004 or Butterworth, 1999) is challenged.

5.2 Recent Studies in Support of the Present Results

Four studies have been published very recently providing converging evidence that subitizing does not occur pre-attentively. Railo et al. (2007) used an inattentional blindness paradigm, in which subjects were surprisingly presented with a collection of dots in the periphery while performing a primary task at centre. Under conditions of inattention (i.e. surprise) no more

than 2 dots could be accurately enumerated. In line with the results from the current study, they showed that enumeration accuracy increased the more attentional resources could be allocated to the enumeration task (i.e. under diverted attention (dual task) and under full attention (single task)). The study bore the confound that the primary task consisted of a length discrimination and thus a magnitude judgement task. Also with respect to the manipulation of attention, Experiment 2 was better controlled than the experiments by Railo et al. (2007). In Experiment 2, subjects were not required to process unexpected (and thus unattended) stimuli, but, due to the load manipulation, were required to process expected stimuli with differential availability of attentional resources. However, Railo et al. (2007) showed that numerosity information is not automatically retrieved from a visual display when task-irrelevant.

Both Olivers and Watson (2008) as well as Egeth, Leonard and Palomares (2008) employed an attentional blink paradigm very similar to the one used by Joseph, Chun and Nakayama (1997). Both presented an RSVP letter identification task together with an enumeration task and measured enumeration performance. Consistent with Experiment 2, both studies found an impairment of subitizing the closer the time lag to preceding letter identification. In contrast to Experiment 2, however, they did not find an effect of reduced attentional resources for numerosity 1. This led Egeth, Leonard and Palomares (2008) to propose that there might be a "magical number 1" instead of a "magical number 4". However, Egeth, Leonard and Palomares (2008) could not replicate Joseph, Chun and Nakayama (1997) in their enumeration condition 0 and 1, and thus also did not replicate the findings from single target detection in Exp. 2.2. This suggests that attentional blink paradigms, and possibly dual-task paradigms in general, might lead to differential results depending on stimulus conditions and task demands.

Interestingly, Olivers and Watson (2008) found that the number of dots also differentially affected the accuracy of the letter identification task. Thus, as in Experiment 2, the primary task also suffers from dual task conditions and, in addition, the number of items can serve as a manipulation of load.

Poiese, Spalek and Di Lollo (2008) approached the question from a slightly different perspective and reasoned that if subitizing was truly pre-attentive, it should be completed in V1 within 50 ms after stimulus onset and before any feedback processing. This hypothesis was

made on the basis of TMS evidence by Corthout et al. (1999) and Ashbridge, Walsh and Cowey (1997). If subitizing is attentive and thus involving areas beyond V1, subitizing accuracy should be higher at higher stimulus presentation times. In line with the findings from Experiment 1, subitizing was better at higher stimulus presentation times and modulated by the number of items (1 and 2). The authors concluded that subitizing must be attentive and involve visual areas beyond V1. The fMRI results of Experiment 3 confirm this conclusion.

5.3   Implications of the Present Results

Both the current results as well as those from the recent literature considerably weaken the claim of both an attentional as well as a functional dichotomy between small and large numerosity judgement. Thus, these results have important implications for the classical theories of visual enumeration that tried to explain such a dichotomy.

5.3.1   Implications for the Theories of Enumeration

Pattern Recognition Account

As mentioned in the General Introduction, the pattern recognition account as originally proposed by Mandler and Shebo (1982) is not an entirely conclusive theory and certainly cannot fully account for the present results. First, the spatial arrangement of stimuli was constantly varied so that the emergence of canonical patterns could only happen occasionally by chance and non-canonical shapes always varied. In order to make a numerosity judgement by recognising a triangle, for example, would require subjects to vary their matching template constantly. Second, pattern recognition predicts a sharp discontinuity between the subitizing and counting ranges which was not observed in either of the employed enumeration paradigms. Third, pattern recognition could only explain the current results if it was equally affected by attentional set and attentional load as enumeration, and thus if it was an attention-demanding process itself. Mandler and Shebo (1982) did not make specific assumptions about the attentional requirements of enumeration or pattern recognition. Fink et al. (2001) showed that deriving either shape or

numerosity from identical visual displays differentially activated attention-related brain areas, suggesting that both operations might draw differentially on attentional processing. Fourth, the strongest argument against pattern recognition might be apparent in the pattern of response variance and Weber fractions: discriminating a dot from a line, or a line from a triangle should be equally effortless according to pattern recognition theory, thus, in the subitizing range, response variance should stay constant and Weber fractions (as adjusted for target number) should decrease. The results from all experiments show that this is clearly not the case and discrimination becomes more difficult with increasing set size also in the subitizing range (as predicted by Weber's law).

Spatial Frequency Account

Given that the experiments by Atkinson et al. (1976 a & b) were one of the first studying subitizing systematically, their results with respect to crowding and perceptual grouping are still very relevant. Experiment 1 demonstrated that, depending on the stimulus properties, subitizing but not detection are compromised and thus perceptual grouping plays a crucial role in enumeration. This is in line with the results of Atkinson, Francis and Campbell (1976). Given that distractor items were spatially intermixed with targets in all cases (and as in Experiment 1, it was unclear which items belonged to the target and which to the distractor set) the attentional effects of the present results, however, cannot be explained by spatial frequency alone.

FINST Account

Pylyshyn's FINST account (Pylyshyn, 1989) states that there is a strict four-item limit of index tokens that are pre-attentively assigned to those items that need to be enumerated or tracked, giving rise to the observed performance discontinuity in enumeration and also multiple object tracking. This hypothesis becomes severely questionable in light of the current findings and those by Joseph, Chun & Nakayama (1997). A single target is well within the capacity limit as suggested by Pylyshyn (1989) and indexing a salient orientation odd-ball (Exp. 1.4 and Joseph, Chun & Nakayama, 1997) or a contrast and orientation odd-ball (Exp. 2.2) should certainly occur pre-attentively. However, the results of Exp. 1.4, Exp. 2.2 and those by Joseph, Chun and

Nakayama (1997) clearly demonstrate that single target detection is affected by reduced availability of attentional resources. Thus, even if indexing itself occurs pre-attentively, some process subsequent to indexing must require attention. Furthermore, the results from Experiment 1 suggest that orientation odd-ball detection and enumeration depend on perceptual grouping and it is unclear how pre-attentive indexing of targets can occur despite possibly attentive perceptual grouping and target-distractor segmentation. In addition, although subitizing accuracy was clearly compromised by the employed attentional manipulations, no clear change in enumeration slopes was observed (though this was not tested formally). It appears that also Trick and Pylyshyn's theory (Trick & Pylyshyn, 1993) suffered from the artificial fitting of bilinear functions to enumeration slopes.

Based on this evidence, it is unlikely that a truly pre-attentive indexing mechanism exists as the basis of subitizing and multiple object tracking and as such a division of attentive and pre-attentive subitizing depending on stimulus properties appears to be rather artificial. The results from this work demonstrate that subitizing is attentive also when stimulus properties induce classical pop-out. In light of the unclear nature of a pre-attentive indexing mechanism, the question is raised whether a true capacity limit of 4 items exist.

Short-term Memory Account

As reviewed in Chapter 1, it is highly controversial whether a fixed-capacity limit of four items exists in either short-term memory or attention (c.f. Bays & Husain, 2008; Davis et al., 2001; Alvarez & Franconeri, 2007). Thus, inferring a fixed set-size limit in enumeration from the proposal of such a set-size limit in visual short-term memory (Cowan, 2001) is problematic.

What is interesting, however, is the striking similarity in performance characteristics between enumeration, visual short-term memory (VSTM) and multiple object-tracking (MOT). In all these processes, performance is very good at a few number of items, and following a sigmoidal function, becomes increasingly worse at higher number of items. The inflection point of the sigmoidal performance functions lies typically between 3 and 5 items, which has led to the controversial interpretation of a fixed set-size limit. The similarity between these differential processes raises the possibility that they share a common capacity-limited process (e.g. Cowan,

2001; Marois & Ivanoff, 2005; Alvarez & Franconeri, 2007), though whether this limit taps memory or attention processes is still debated (c.f. Fougnie & Marois, 2006). For example, a multiple object tracking task and visual short-term memory task interfere with each other (Fougnie & Marois, 2006) and activate partially overlapping brain areas in posterior parietal cortex as suggested by Marois & Ivanoff (2005) in a meta-analysis. However, a possible link between VSTM and MOT (or even enumeration) on a neural level has never been investigated. The results of Experiment 3, however, together with the findings from Todd, Fougnie and Marois (2005) suggest that both enumeration and visual short-term memory may share a common processing site in temporo-parietal junction which is modulated by the number of items enumerated or memorised.

Thus, although the current results do not support the idea of a strict set-size limit at 4 items, it might well be possible that enumeration and visual short-term memory share an underlying process characterised by the sigmoidal performance function (thus, by a capacity limit, but not a strict set-size limit). It might even be possible that enumeration relies on visual short-term memory and thus the similarity in performance characteristics arises entirely from the capacity limit in VSTM. In fact, a possible link between enumeration and VSTM has never been studied.

If it is true that enumeration depends on visual short-term memory and is constrained by the same capacity limit, the criticism raised in Experiment 1 that cueing interferes with memory and not with attention might be justified. It is however, not the delay between stimulus display and response that leads to a decay in memory and thus less accurate response (as discussed in section 2.6.2), but the fact that memory capacity is reached when two subsets have to be enumerated in parallel. Thus, it is the visual-spatial aspect of VSTM that might be limiting here, not the temporal aspect. It is therefore possible that post-cueing leads to impaired enumeration accuracy because encoding all items of two subsets exceeds the capacity limit of VSTM (except for the rare cases when both subsets are very small). However, this possibility can only occur if encoding into VSTM is done by item rather then by numerosity as a feature. As argued in section 2.6.2, if numerosity is a feature as colour or shape and as such encoded into VSTM, memorising the numerosity of two subsets in parallel should not be limiting. If, however, every item takes up a "slot" in VSTM, then the effect of post-cueing might have been

due to memory overload.

Two arguments can be made here. First, it is highly controversial in the VSTM literature what counts as "slot" in VSTM, and whether it is the number of features or the number of objects that constitute the capacity limit (e.g. Olson & Jiang, 2002; Davis & Holmes, 2005). As mentioned in Chapter 1, the question of what counts as an "item" and deserves the assignment of a "token" is also problematic in the FINST account (Pylyshyn, 1989). Thus, it is unclear whether the dots in Experiment 1 were encoded as individual objects or whether the set numerosity was encoded as "numerosity feature". Second, even if subitizing is limited by visual-short term memory, this does not affect the original hypothesis of this thesis. It is assumed that encoding into short-term memory depends on attention (Awh, Jonides & Reuter-Lorenz, 1998; Cowan, 2001), thus even if the retrieval of numerosity from a display may occur pre-attentively, its encoding into VSTM is attentive. Thus, if subitizing depends on short-term memory, it must be attentive. Therefore, although it cannot be entirely excluded that the cueing manipulation of Experiment 1 may have tapped into a memory process, it must have involved attention.

5.3.2   Implications for Other Aspects of the Literature

The imaging results are interesting with respect to the claim of preserved subitizing abilities in neglect patients (Vuilleumier & Rafal, 2000). The ventral fronto-parietal network, and thus temporo-parietal junction, is, according to Corbetta and Shulman (2002) one of the main sites of damage in hemispatial neglect (although there is some controversy about this, others claim that more dorsal regions are responsible for the neglect phenomenon (Mesulam, 1999)). In fact, three of the four tested patients in the study by Vuilleumier and Rafal (2002) showed damage in right TPJ (as well as widespread damage in other right parietal areas). Given that stimulus-driven attention is impaired in neglect patients and that right TPJ is modulated by the number of items attended, it might be possible that the failure to distinguish 1 from 2 items in the neglected field (as suggested by the behavioural data of Vuilleumier and Rafal (2002)) is indeed due to an inability to subitize. As raised in Chapter 1, this points to the possibility that in the cases where patients correctly "enumerated" four items (distributed across both hemifields), they in fact made a categorical decision rather than a true subitizing operation involving stimulus-driven

attention.

As Di Lollo et al. (2001) have argued in the case of visual search, the results from this work demonstrate that the inference of attentional requirements from performance slopes is flawed also in enumeration. Furthermore, the hypothesis of a bimodal model for enumeration is as problematic as it is for visual search. The proposal by Norman and Bobrow (1975) provides a good account for the allocation of resources also in the case of enumeration and could reconcile the findings from older studies with the present one. Under traditional testing conditions (unlimited presentation times, no distractors) subitizing is a very easy task and only very few attentional resources are required to reach maximal performance. If subitizing is made harder either by stimulus properties (short stimulus exposure, presence of distractors, absence of covarying parameters, visual complexity of stimuli) or by imposing another task, processing resources need to be shared out amongst enumeration, processes on which enumeration is dependent on (target - distractor segregation, perceptual grouping) and additional task demands. If the sum of all these processes needs more resources than are available, this leads to a failure to reach maximal performance in either task, i.e. impaired enumeration and dual task costs also in the additional task. This account implies that attentional effects on subitizing can only be seen if it is rendered difficult and resources need to be shared out between other processes. Thus, previous studies might have failed to see attentional effects on subitizing due to the employed testing conditions. Counting however, requires more resources due to the increased amount of information to be processed and thus maximum resources are exhausted much earlier - and in a manner clearly modulated by the number of items presented. This has led to the classical interpretation that counting is attentive as the resource limit is obvious already under traditional testing conditions. The implication of Norman and Bobrow's account is that resources are limited and need to be shared on all task demands at hand, thus the balance of resource allocation between tasks is the crucial aspect in attentional processing and not the question of whether there are tasks that are independent of resource allocation. In this respect, the account is very similar to Lavie's load theory: the balance between processing resources is shifted between tasks according to their perceptual requirements, i.e. the load of each task (Lavie, 2005). Thus, the results of this study can be reconciled with previous studies in that the balance of attentional resources was biased differentially towards numerosity processing.

For future research, it might thus be more interesting to study how attentional resources are shared between the different processes that are involved in enumeration and which process is the most attention demanding component.

5.4   A Tentative Proposal

Based on the results of this work and others reported in the literature, the following view on visual enumeration is proposed. Subitizing might not be a very "special" or "magical" process, but in fact the reflection of easy, super-threshold and thus effortless estimation. Under stimulus conditions making subitizing very easy (as employed in many previous studies), the data pattern might look as if subitizing performance is indeed exceptionally good. The evidence in this work however shows that if subitizing is made difficult, it can be compromised to the same extent as the enumeration of larger numerosities. As no differential effects of attentional withdrawal between small and large numerosity judgement was observed, these results suggest that enumeration is subserved by a single, continuous enumeration mechanism. This mechanism is equivalent to estimation and follows Weber's law. As a consequence of Weber's law, subitizing is nothing more than effortless and accurate estimation due to the little amount of information that needs to be handled, in accordance with Ross (2003). At higher numerosities, estimation becomes increasingly inaccurate and in order to overcome this inaccuracy, a learned verbal counting mechanism has evolved. Thus, it is proposed that there is no dichotomy between small and large numerosity judgement *per se*, but, within the higher numerosities, there is a dichotomy between fast and increasingly inaccurate estimation and accurate, but increasingly slow verbal counting. This is in line with the proposal by Gallistel and Gelman (2002).

However, whether subitizing and estimation are truly subserved by the same neural mechanism has not been formally investigated in this thesis and further studies are required to address this issue.

It should be mentioned that Revkin et al. (2008) very recently investigated the question whether subitizing and estimation reflect the same cognitive process. They compared the enumeration of 1 to 8 dots with enumeration of 10 to 80 dots (without distractors) and mea-

sured the variation coefficient of response (which is similar to the measure of response variance employed in the current work). Based on the fact that response variability was very low in the subitizing range (with a clear discontinuity towards the higher numerosities) and much higher in the range from 10 to 40, they propose that subitizing must be a distinct process from estimation and cannot be explained by Weber's law. However, the authors neglect the argument made by Ross (2003): that discriminating up to four items is a super-threshold task and as such, not much response variability should be expected. The fact that variability in the 10 - 40 range was much higher is probably due to the fact that much more visual information had to be processed, leading to increased noise and thus, greater response variability. Particularly as Revkin et al. (2008) did not use distractors, noise was absent in the subitizing range, so the ceiling effect of no variability in this range is not surprising. That is, this design again suffered from an imbalance in processing effort and perceptual noise between the compared conditions. It appears that only if small and large numerosities are equated for task difficulty and noise levels, a direct comparison would be meaningful.

Nieder & Merten (2007), however, provided neurophysiological evidence for a single, continuous enumeration mechanism. They demonstrated that tuning functions of numerosity-selective neurons in monkey pre-frontal cortex exhibit a clear Weber-fraction signature for numerosities from 1 to 30. Weber fractions were constant across the whole numerosity range and did not show any discontinuity in the subitizing range.

In line with these finding, Burr and Ross (2008) recently showed in a very elegant adaptation study that adapting to small numerosities makes an array of large numerosities appear smaller and vice versa, adapting to large numerosities makes an array of small numerosities appear larger. They revealed this effect on a wide range of numerosities suggesting that there is a single neural mechanism underlying both small and large numerosity perception. However, they did not test many numerosities in the range below 10, so an inference for very small numerosities is difficult to make.

Clearly, this issue is not resolved yet and further studies are needed. An approach similar to that from Burr and Ross (2008) to investigate the proposal of a single enumeration mechanism also in small numerosity ranges is proposed in the following section.

5.5   Future Research

There are several possibilities to extend the experiments presented here or to address some of their weaknesses. Furthermore, the results of this work point to new aspects of enumeration that may be worth investigating.

With regard to the paradigm of Experiment 1, the attentional effect of the post-cue condition is slightly ambiguous as it is is unclear whether subjects spread their attention equally towards both subsets or whether they randomly chose one subset to attend to. A more balanced approach would be to introduce both a pre-cue and a post-cue in every trial and make them either congruent or incongruent in unequal proportions. The pre-cue directs attention towards a certain subset, whereas the post-cue indicates the subset whose numerosity shall be reported. For example, in 80% of the trials, both pre- and post-cue are identical and therefore induce a clear attentional advantage towards the cued subset. In about 20% of the trials, the cues are incongruent such that subjects have to report the unattended subset. With this approach, the line between an attended and an unattended situation could be drawn much clearer than with the current experimental design. Furthermore, the memory confound might be eliminated as there is a post-cue in all trials.

In order to investigate further the role of perceptual grouping in enumeration, a design varying both the ease of perceptual grouping as well as the attentional demands would be an interesting approach. For example, a 2 by 2 design could be employed, making perceptual grouping either easy or difficult and at the same time, varying the amount of attention available to the task (e.g. attended versus unattended). Such a design could reveal how attention and perceptual grouping interact in enumeration.

With respect to the paradigm of Experiment 2, it would be interesting to investigate how approximate numerosity judgement is affected by dual task and attentional load. It might be conceivable that an approximate judgement, i.e. the distinction between "rather few" or "rather more" items might be part of gist perception or scene scrutinisation and thus might survive dual task conditions (Li et al., 2002), or at least be less susceptible for manipulations of attentional load. This would reveal whether numerosity judgement could be divided into a rough

approximation step that does not require much attention or processing resources and an exact and discrete enumeration step that requires attention and is needed to accurately discriminate adjacent numerosities from each other.

Furthermore, the fMRI results of Experiment 3 can be analysed in many different ways. One approach would be to identify those areas in the brain whose neural activations are parametrically modulated by numerosity and to analyse how attentional load affects the activation levels in these areas. Similarly, activations that are parametrically modulated by a load-induced change in Weber fraction may point to those areas that reflect the ability to discriminate numerosities from each other.

In order to investigate the idea of a single numerosity judgement mechanism across low and high numerosities, a behavioural adaptation paradigm similar to the one by Burr and Ross (2008) could be developed. Numerosities within and outside the subitizing range could be used and processing effort equated by using distractors. Subjects could be adapted to numerosities either below or above 5 and then asked to judge the numerosity of a neutral test condition (numerosity 5). Following the logic of behavioural adaptation, subjects should systematically overestimate the numerosity of the neutral condition if they have adapted previously to numerosities below 5. If they have adapted to numerosities above 5 they should systematically underestimate the numerosity of the neutral condition. If these predictions are fulfilled, then this would allow the conclusion that numerosity judgements in both ranges are based on the same neural substrate and share a common cognitive processing mechanism.

5.6   Final Remarks

In his paper from 1871, Stanley Jevons concluded:

"My conclusion that the number five is beyond the limit of perfect discrimination, by some persons at least, is strongly supported by the principles of rhythm. All the kinds of time employed by musicians depend upon a division of the bar into two or three equal parts, or into multiple of these. [...] Quinary music, even if it could be executed, would be ill appreciated by the hearers, and, though all the powers of the human mind may be expected to progress in the

course of ages, quinary rhythm belongs to the music of the distant future."

       Now, more than 130 years later, quinary rhythm in music is something rather common and well appreciated by some hearers. Although the powers of the human mind might not have progressed as far as to appreciate more than five items at a glance, the idea of a fixed limit of discrimination needs to be reconsidered when thinking about how many items the mind is able to embrace.

References

Akin, O. & Chase, W. Quantification of three-dimensional structures. *J Exp Psychol Hum Percept Perform* **4**, 397-410 (1978).

Alston, L. & Humphreys, G. W. Subitization and attentional engagement by transient stimuli. *Spat Vis* **17**, 17-50 (2004).

Alvarez, G. A. & Cavanagh, P. The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychol Sci* **15**, 106-11 (2004).

Alvarez, G. A. & Franconeri, S. L. How many objects can you track?: Evidence for a resource-limited attentive tracking mechanism. *J Vision* **7**, 1-10 (2007).

Ansari, D., Lyons, I. M., van Eimeren, L. & Xu, F. Linking visual attention and number processing in the brain: the role of the temporo-parietal junction in small and large symbolic and nonsymbolic number comparison. *J Cogn Neurosci* **19**, 1845-53 (2007).

Ashbridge, E., Walsh, V. & Cowey, A. Temporal aspects of visual search studied by transcranial magnetic stimulation. *Neuropsychologia* **35**, 1121-31 (1997).

Atkinson, J., Campbell, F. W. & Francis, M. R. The magic number 4 +/- 0: a new look at visual numerosity judgements. *Perception* **5**, 327-34 (1976a).

Atkinson, J., Francis, M. R. & Campbell, F. W. The dependence of the visual numerosity limit on orientation, colour, and grouping in the stimulus. *Perception* **5**, 335-42 (1976b).

Averbach, E. The span of apprehension as a function of exposure duration. *J Verb Learn Verb Behav* **2**, 60-64 (1963).

Awh, E., Jonides, J. & Reuter-Lorenz, P. A. Rehearsal in spatial working memory. *J Exp Psychol Hum Percept Perform* **24**, 780-90 (1998).

Balakrishnan, J. D. & Ashby, F. G. Is subitizing a unique numerical ability? *Percept Psychophys*, **50**, 555-64 (1991).

Balakrishnan, J. D. & Ashby, F. G. Subitizing: magical numbers or mere superstition? *Psychol Res*, **54**, 80-90 (1992).

Bays, P. M. & Husain, M. Dynamic shifts of limited working memory resources in human vision. *Science* **321**, 851-4 (2008).

Ben-Av, M. B., Sagi, D. & Braun, J. Visual attention and perceptual grouping. *Percept Psychophys* **52**, 277-94 (1992).

Bourdon, B. Sur le temps necessaire pour nommer les nombres. *Revue philosophique de la France et de l'etranger* **65**, 426-431 (1908).

Brannon, E. M. & Terrace, H. S. Representation of the numerosities 1-9 by rhesus macaques (Macaca mulatta). *J Exp Psychol Anim Behav Process* **26**, 31-49 (2000).

Braun, J. Shape-from-shading is independent of visual attention and may be a 'texton'. *Spat Vis* **7**, 311-22 (1993).

Braun, J. Vision and attention: the role of training. *Nature* **393**, 424-5 (1998).

Burr, D. & Ross, J. A visual sense of number. *Curr Biol* **18**, 425-8 (2008).

Bush, G., Whalen, P. J., Rosen, B. R., Jenike, M. A., McInerney, S. C. & Rauch , S. L. The counting Stroop: an interference task specialized for functional neuroimaging–validation study with functional MRI. *Hum Brain Mapp* **6**, 270-82 (1998).

Butterworth, B. The Mathematical Brain. *Macmillan*, London (1999).

Castelli, F., Glaser, D. E. & Butterworth, B. Discrete and analogue quantity processing in the parietal lobe: a functional MRI study. *Proc Natl Acad Sci USA* **103**, 4693-8 (2006).

Cipolotti, L., Butterworth, B. & Denes, G. A specific deficit for numbers in a case of dense acalculia. *Brain* **114** (Pt 6), 2619-37 (1991).

Corbetta, M., Kincade, J. M., Ollinger, J. M., McAvoy, M. P. & Shulman, G. L. Voluntary orienting is dissociated from target detection in human posterior parietal cortex. *Nat Neurosci* **3**, 292-7 (2000).

Corbetta, M., Kincade, J. M. & Shulman, G. L. Neural systems for visual orienting and their relationships to spatial working memory. *J Cogn Neurosci* **14**, 508-23 (2002).

Corbetta, M. & Shulman, G. L. Control of goal-directed and stimulus-driven attention in the brain. *Nat Rev Neurosci* **3**, 201-15 (2002).

Cordes, S., Gelman, R., Gallistel, C. R. & Whalen, J. Variability signatures distinguish verbal from nonverbal counting for both large and small numbers. *Psychon Bull Rev* **8**, 698-707 (2001).

Corthout, E., Uttl, B., Walsh, V., Hallett, M. & Cowey, A. Timing of activity in early visual cortex as revealed by transcranial magnetic stimulation. *Neuroreport* **10**, 2631-4 (1999).

Cowan, N. The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behav Brain Sci* **24**, 87-114; discussion 114-85 (2001).

Davis, G. & Holmes, A. The capacity of visual short-term memory is not a fixed number of objects. *Mem Cognit* **33**, 185-95 (2005).

Davis, G., Welch, V. L., Holmes, A. & Shepherd, A. Can attention select only a fixed number of objects at a time? *Perception* **30**, 1227-48 (2001).

Dehaene, S. The Number Sense: How the Mind Creates Mathematics. *Oxford University Press*, New York (1997).

Dehaene, S. & Changeux, J. P. Development of elementary numerical abilities: A neuronal model. *J Cogn Neurosci* **5**, 390-407 (1993).

Dehaene, S. & Cohen, L. Dissociable mechanisms of subitizing and counting: neuropsychological evidence from simultanagnosic patients. *J Exp Psychol Hum Percept Perform*, **20**, 958-75 (1994).

Dehaene, S., Piazza, M., Pinel, P. & Cohen, L. Three parietal circuits for number processing. *Cogn Neuropsychol* **20**, 487-506 (2003).

Di Lollo, V., Kawahara, J., Zuvic, S. M. & Visser, T. A. The preattentive emperor has no clothes: a dynamic redressing. *J Exp Psychol Gen* **130**, 479-92 (2001).

Dojat, M., Piettre L., Delon-Martin, C., Pachot-Clouard, M., Segebarth, C. & Knoblauch, K. Global integration of local color differences in transparency perception: An fMRI study. *Vis Neurosci* **23**, 357-64 (2006).

Driver, J., Davis, G., Russell, C., Turatto, M. & Freeman, E. Segmentation, attention and phenomenal visual objects. *Cognition* **80**, 61-95 (2001).

Duncan, J. & Humphreys, G. W. Visual search and stimulus similarity. *Psychol Rev*, **96**, 433-58 (1989).

Egeth, H. E., Leonard, C. L. & Palomares, M. The role of attention in subitizing: Is the magical number 1? *Vis Cogn* **16**, 463-473 (2008).

Feigenson, L., Dehaene, S. & Spelke, E. Core systems of number. *Trends Cogn Sci* **8**, 307-14 (2004).

Fink, G. R., Marshall, J. C., Gurd, J., Weiss, P. H., Zafiris, O., Shah, N. J. & Zilles, K. Deriving numerosity and shape from identical visual displays. *Neuroimage* **13**, 46-55 (2001).

Folk, C. L., Egeth, H. & Kwak, H. W. Subitizing: direct apprehension or serial processing? *Percept Psychophys*, **44**, 313-20 (1988).

Forstmann, B. U., Wolfensteller, U., Derrfuss, J., Neumann, J., Brass, M., Ridderinkhof, K. R. & von Cramon, D. Y. When the choice is ours: context and agency modulate the neural bases of decision-making. *PLoS ONE* **3**, e1899 (2008).

Fougnie, D. & Marois, R. Distinct capacity limits for attention and working memory: Evidence from attentive tracking and visual working memory paradigms. *Psychol Sci* **17**, 526-34 (2006).

Found, A. & Muller, H. J. Searching for unknown feature targets on more than one dimension: investigating a "dimension-weighting" account. *Percept Psychophys* **58**, 88-101 (1996).

Franconeri, S. L., Alvarez, G. A. & Enns, J. T. How many locations can be selected at once? *J Exp Psychol Hum Percept Perform* **33**, 1003-12 (2007).

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J. B., Frith, C. D. & Frackowiak, R. S. J. Statistical parametric maps in functional imaging: a general linear approach. *Human Brain Mapp* **2**, 189-210 (1995).

Gallistel, C. R. & Gelman, R. Non-verbal numerical cognition: from reals to integers. *Trends Cogn Sci* **4**, 59-65 (2000).

Halberda, J., Sires, S. F. & Feigenson, L. Multiple spatially overlapping sets can be enumerated in parallel. *Psychol Sci* **17**, 572-6 (2006).

Hamilton, W. The metaphysics of Sir Wiliam Hamilton (ed. Bowen, F.) *Sever & Francis*, Cambridge (1866).

Hauser, M. D., Carey, S. & Hauser, L. B. Spontaneous number representation in semi-free-ranging rhesus monkeys. *Proc Biol Sci* **267**, 829-33 (2000).

Jensen, E. M., Reese, E. P. & Reese, T. W. The subitizing and counting of visually presented fields of dots. *J Psychol* **30**, 363-92 (1950).

Jevons, W. S. The Power of Numerical Discrimination. *Nature* **3**, 281-282 (1871).

Joseph, J. S., Chun, M. M. & Nakayama, K. Attentional requirements in a 'preattentive' feature search task. *Nature* **387**, 805-7 (1997).

Joseph, J. S., Chun, M. M. & Nakayama, K. Reply to Braun: Vision and attention: the role of training. *Nature* **393**, 425 (1998).

Julesz, B. Towards an axiomatic theory of pre-attentive vision. (eds. Edelman, G., Gall, W. & Cowan, W.) *Wiley*, Toronto, Ontario, Canada (1984).

Karnath, H. O., Ferber, S. & Himmelbach, M. Spatial awareness is a function of the temporal not the posterior parietal lobe. *Nature* **411**, 950-3 (2001).

Kaufman, E. L. & Lord, M. W. The discrimination of visual number. *Am J Psychol*, **62**, 498-525 (1949).

Kawabe, T. & Miura, K. Perceptual grouping in shape from shading. *Perception* **33**, 601-14 (2004).

Kincade, J. M., Abrams, R. A., Astafiev, S. V., Shulman, G. L. & Corbetta, M. An event-related functional magnetic resonance imaging study of voluntary and stimulus-driven orienting of attention. *J Neurosci* **25**, 4593-604 (2005).

Klahr, D. Quantification Processes. (ed. Chase, W. G.) *Academic Press* (1973).

Kleffner, D. A. & Ramachandran, V. S. On the perception of shape from shading. *Percept Psychophys* **52**, 18-36 (1992).

Lamy, D., Segal, H. & Ruderman, L. Grouping does not require attention. *Percept Psychophys* **68**, 17-31 (2006).

Lavie, N. Perceptual load as a necessary condition for selective attention. *J Exp Psychol Hum Percept Perform* **21**, 451-68 (1995).

Lavie, N. Distracted and confused? Selective attention under load. *Trends Cogn Sci* **9**, 75-82 (2005).

Li, F. F., VanRullen, R., Koch, C. & Perona, P. Rapid natural scene categorization in the near absence of attention. *Proc Natl Acad Sci U S A* **99**, 9596-601 (2002).

Luck, S. J. & Vogel, E. K. The capacity of visual working memory for features and conjunctions. *Nature* **390**, 279-81 (1997).

Mack, A., Tang, B., Tuma, R., Kahn, S. & Rock, I. Perceptual organization and attention. *Cognit Psychol* **24**, 475-501 (1992).

Maguire, E. A., Frith, C. D., Burgess, N., Donnett, J. G. & O'Keefe, J. Knowing where things are: parahippocampal involvement in encoding object locations in virtual large-scale space. *J Cogn Neurosci* **10**, 61-76 (1998).

Mandler, G. & Shebo, B. J. Subitizing: an analysis of its component processes. *J Exp Psychol Gen* **111**, 1-22 (1982).

Marois, R. & Ivanoff, J. Capacity limits of information processing in the brain. *Trends Cogn Sci* **9**, 296-305 (2005).

Marois, R., Leung, H. C. & Gore, J. C. A stimulus-driven approach to object identity and location processing in the human brain. *Neuron* **25**, 717-28 (2000).

Mesulam, M. M. Spatial attention and neglect: parietal, frontal and cingulate contributions to the mental representation and attentional targeting of salient extrapersonal events. *Philos Trans R Soc Lond B Biol Sci* **354**, 1325-46 (1999).

Moore, C. M. & Egeth, H. Perception without attention: evidence of grouping under conditions of inattention. *J Exp Psychol Hum Percept Perform* **23**, 339-52 (1997).

Muller, H. J., Heller, D. & Ziegler, J. Visual search for singleton feature targets within and across feature dimensions. *Percept Psychophys* **57**, 1-17 (1995).

Nakayama, K. & Joseph, J. S. Attention, pattern recognition, and pop-out in visual search. (ed. Parasuraman, R.) *MIT Press, Cambridge*, MA (1998).

Nakayama, K. & Mackeben, M. Sustained and transient components of focal visual attention. *Vision Res* **29**, 1631-47 (1989).

Nieder, A., Freedman, D. J. & Miller, E. K. Representation of the quantity of visual items in the primate prefrontal cortex. *Science* **297**, 1708-11 (2002).

Nieder, A. & Merten, K. A labeled-line code for small and large numerosities in the monkey prefrontal cortex. *J Neurosci* **27**, 5986-93 (2007).

Nieder, A. & Miller, E. K. Analog numerical representations in rhesus monkeys: evidence for parallel processing. *J Cogn Neurosci*, **16**, 889-901 (2004).

Norman, D. A. & Bobrow, D. G. On data-limited and resource-limited processes. *Cogn Psychol* **7**, 44-64 (1975).

O'Connor, D. H., Fukui, M. M., Pinsk, M. A. & Kastner, S. Attention modulates responses in the human lateral geniculate nucleus. *Nat Neurosci* **5**, 1203-9 (2002).

Olivers, C. N. L. & Watson, D. G. Subitizing requires attention. *Vis Cogn* **16**, 439-462 (2008).

Olson, I. R. & Jiang, Y. Is visual short-term memory object based? Rejection of the "strong-object" hypothesis. *Percept Psychophys* **64**, 1055-67 (2002).

Piazza, M., Giacomini, E., Le Bihan, D. & Dehaene, S. Single-trial classification of parallel pre-attentive and serial attentive processes using functional magnetic resonance imaging. *Proc R Soc Lond B Biol Sci*, **270**, 1237-45 (2003).

Piazza, M., Izard, V., Pinel, P., Le Bihan, D. & Dehaene, S. Tuning curves for approximate numerosity in the human intraparietal sulcus. *Neuron* **44**, 547-55 (2004).

Piazza, M., Mechelli, A., Butterworth, B. & Price, C. J. Are subitizing and counting implemented as separate or functionally overlapping processes? *Neuroimage*, **15**, 435-46 (2002).

Poiese, P., Spalek, T. M. & Di Lollo, V. Attentional involvement in subitizing: Questioning the preattentive hypothesis. *Vis Cogn* **16**, 474-485 (2008).

Puts, M. J. H. & de Weert, C. M. M. Does color influence subitization? *Acta Psychol* **97**, 71-78 (1997).

Pylyshyn, Z. The role of location indexes in spatial perception: a sketch of the FINST spatial-index model. *Cognition*, **32**, 65-97 (1989).

Railo, H., Koivisto, M., Revonsuo, A. & Hannula, M. M. The role of attention in subitizing. *Cognition* **107**, 82-104 (2008).

Ramachandran, V. S. Perceiving shape from shading. *Sci American* **256**, 76-83 (1988).

Raymond, J. E., Shapiro, K. L. & Arnell, K. M. Temporary suppression of visual processing in an RSVP task: an attentional blink? *J Exp Psychol Hum Percept Perform* **18**, 849-60 (1992).

Revkin, S. K., Piazza, M., Izard, V., Cohen, L. & Dehaene, S. Does subitizing reflect numerical estimation? *Psychol Sci* **19**, 607-14 (2008).

Ross, J. Visual discrimination of number without counting. *Perception* **32**, 867-70 (2003).

Sagi, D. & Julesz, B. Fast noninertial shifts of attention. *Spat Vis*, **1**, 141-9 (1985).

Sathian, K., Simon, T. J., Peterson, S., Patel, G. A., Hoffman, J. M. & Grafton, S. T. Neural evidence linking visual object enumeration and attention. *J Cogn Neurosci*, **11**, 36-51 (1999).

Schwartz, S., Vuilleumier, P., Hutton, C., Maravita, A., Dolan, R. J. & Driver, J. Attentional load and sensory competition in human vision: modulation of fMRI responses by load at fixation during task-irrelevant stimulation in the peripheral visual field. *Cereb Cortex* **15**, 770-86 (2005).

Serences, J. T., Shomstein, S., Leber, A. B., Golay, X., Egeth, H. E. & Yantis, S. Coordination of voluntary and stimulus-driven attentional control in human cortex. *Psychol Sci* **16**, 114-22 (2005).

Simon, T. J. & Vaishnavi, S. Subitizing and counting depend on different attentional mechanisms: evidence from visual enumeration in afterimages. *Percept Psychophys*, **58**, 915-26 (1996).

Sperling, G. The information available in brief visual presentations. *Psychological Monographs*, **74** (1960).

Starkey, P. & Cooper, R. G., Jr. Perception of numbers by human infants. *Science* **210**, 1033-5 (1980).

Taves, E. H. Two mechanisms for the perception of visual numerousness. *Arch Psychol* **37**, 1-47 (1941).

Todd, J. J., Fougnie, D. & Marois, R. Visual short-term memory load suppresses temporo-parietal junction activity and induces inattentional blindness. *Psychol Sci* **16**, 965-72 (2005).

Todd, J. J. & Marois, R. Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature* **428**, 751-4 (2004).

Treisman, A. & Gormican, S. Feature analysis in early vision: evidence from search asymmetries. *Psychol Rev* **95**, 15-48 (1988).

Treisman, A. M. & Gelade, G. A feature-integration theory of attention. *Cognit Psychol*, **12**, 97-136 (1980).

Trick, L. M. & Enns, J. T. Measuring preattentive processes: When is pop-out not enough? *Vis Cogn* **4**, 163-198 (1997).

Trick, L. M. & Pylyshyn, Z. W. What enumeration studies can show us about spatial attention: evidence for limited capacity preattentive processing. *J Exp Psychol Hum Percept Perform*, **19**, 331-51 (1993).

Trick, L. M. & Pylyshyn, Z. W. Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. *Psychol Rev*, **101**, 80-102 (1994).

Van Oeffelen, M. P. & Vos, P. G. A probabilistic model for the discrimination of visual number. *Percept Psychophys* **32**, 163-170 (1982).

Vetter, P., Butterworth, B. & Bahrami, B. Modulating attentional load affects numerosity estimation: evidence against a pre-attentive subitizing mechanism. *PLoS ONE*, **3(9)**, e3269 (2008).

Vogel, E. K., Woodman, G. F. & Luck, S. J. Storage of features, conjunctions and objects in visual working memory. *J Exp Psychol Hum Percept Perform* **27**, 92-114 (2001).

Vuilleumier, P. O. & Rafal, R. D. A systematic study of visual extinction. Between- and within-field deficits of attention in hemispatial neglect. *Brain*, **123** (Pt 6), 1263-79 (2000).

Wang, Q., Cavanagh, P. & Green, M. Familiarity and pop-out in visual search. *Percept Psychophys*, **56**, 495-500 (1994).

Watson, D. G. & Maylor, E. A. Effects of color heterogeneity on subitization. *Percept Psychophys* **68**, 319-26 (2006).

Wender, K. F. & Rothkegel, R. Subitizing and its subprocesses. *Psychol Res* **64**, 81-92 (2000).

Wolfe, J. M. Guided search 2.0: A revised model of visual search. *Psychon Bull Rev* **1**, 202-238 (1994).

Wolfe, J. M. Visual memory: what do you know about what you saw? *Curr Biol* **8**, R303-4 (1998).

Xu, F. Numerosity discrimination in infants: evidence for two systems of representations. *Cognition* **89**, B15-25 (2003).

Zhang, W. & Luck, S. J. Discrete fixed-resolution representations in visual working memory. *Nature* **453**, 233-5 (2008).