ORIGINAL ARTICLE

Matching of visual input to only one item at any one time

Roos Houtkamp · Pieter R. Roelfsema

Abstract When we perform a visual search we know what we are looking for and determine where it is. A representation of the object in our working memory, the 'search-template', is compared to the items in the scene until a match is found. So far it is unknown whether observers can search for multiple items at the same time. Here we compare the performance of subjects between a task in which they search for one of two target-items in a stream of visual objects and a task with only a single target. We find that search is effectively limited to one item at a time. This limitation occurs for simple and complex objects and even if the subjects have to look for two features from different domains. We conclude that matching has a fundamental capacity-limitation as the visual input can be matched to only one search-template at a time.

R. Houtkamp · P. R. Roelfsema
Department of Vision and Cognition,
Netherlands Institute for Neuroscience,
an Institute of the Royal Netherlands Academy of Arts and
Sciences (KNAW), Meibergdreef 47, 1105 BA Amsterdam,
The Netherlands

P. R. Roelfsema
Department of Experimental Neurophysiology,
Center for Neurogenomics and Cognitive Research,
VU University, De Boelelaan 1085, 1081 HV Amsterdam,
The Netherlands

R. Houtkamp (✉)
Department of Cognitive Biology, Otto-von-Guericke
Universitaet, Leipzigerstrasse 44, Haus 91,
39120 Magdeburg, Germany
e-mail: rozemarijn.houtkamp@nat.uni-magdeburg.de

Introduction

We start the day by searching for the button to silence the alarm clock and end the day looking for our toothpaste. This process of visual search is an essential ingredient of almost all activities of our daily life that entails a tight interaction between working memory and the representation of the visual scene. During visual search a representation of the object that we are looking for, the 'search-template', has to be maintained in short-term memory and to be compared to the incoming visual information until a match is found (Desimone & Duncan, 1995; Wolfe, 1994). Previous research demonstrated that working memory can hold approximately 3–4 items (Cowan, 2001; Luck & Vogel, 1997). Here we ask whether multiple items in working memory can be matched in parallel to the incoming visual information.
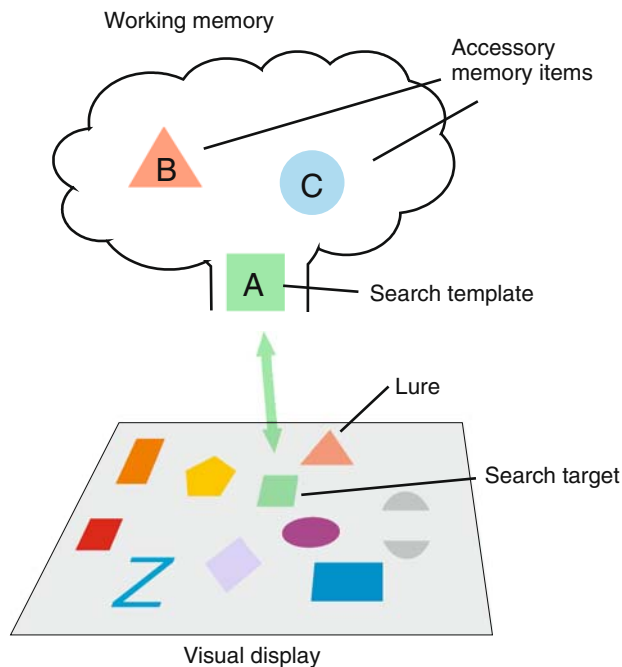
Recent studies started to investigate the relationship between working memory and visual search by examining the influence of extra, 'accessory' items in working memory. Subjects looked for item A, while they stored items B and C in memory for a later task (Fig. 1). If all items in working memory would have a similar status, then distractor items of type B or C in the visual display (lures) should cause more interference during the search for A than other distractors, because lures match an item in memory. It is generally found that subjects are quite accurate in such a task and that the lures cause few false alarms (Downing & Dodds, 2004; Houtkamp & Roelfsema, 2006; Olivers, Meijer & Theeuwes 2006; Soto, Heinke, Humphreys & Blanco 2005), but the aforementioned studies do not fully agree on the amount of residual control that is exerted by the accessory items. Some studies did not observe interference (Downing & Dodds, 2004; Houtkamp & Roelfsema, 2006), while other studies did find interference (Olivers et al., 2006; Soto et al., 2005)

Fig. 1 Organization of working memory during a search task. The search-template is stored in short-term memory and matched to the display items until the target is detected. Display-items that match one of the accessory items in short-term memory (lures) should not be detected

or even facilitation, i.e. lures were rejected faster than regular distractor items during search (Woodman & Luck, 2007).

The differences among findings may depend on whether the target representation occupied space in working memory. In the studies that did not find interference, the subjects had to actively memorize the current search-template, because it changed from trial to trial (variable mapping in the terminology of Schneider & Shiffrin, 1977). Such an active memory representation of the search-template seems to block the access of the accessory memory-items to the visual representation (bottleneck in Fig. 1). In the studies that found interference or facilitation, the target was an item that differed from all the other items (pop-out search, Olivers et al., 2006), or it remained the same across many trials (consistent mapping; Soto et al., 2005; Woodman & Luck, 2007; Schneider & Shiffrin, 1977) so that the target representation occupied little or no space in working memory. Findings of Oh and Kim (2003) and Olivers (2008), who directly compared searches for items that did and did not occupy space in working memory, confirm this interpretation. Accessory memory items interfered if the search target did not occupy space in working memory, but did not if subjects had to memorize a new target on every trial, in accordance with the bottleneck model of Fig. 1. These results, taken together, imply that the search-template occupies a special slot in working

memory that, when filled, prevents other memory-items from accessing the visual representation. If this slot is not filled, residual interference or facilitation by the accessory memory-items can occur.

At this point it is tempting to conclude that the special slot in working memory can only hold one active search-template at a time, although more than one item can be stored in memory. However, this conclusion may be premature, because it was to the subjects' strategic advantage in the previous studies to keep the accessory memory-items in a passive state as this would prevent interference. What happens to the distinction between search-template and accessory items if the task demands multiple search-templates to be active in parallel? To address this question, we will now *require* the subjects to maintain two active search-templates, asking them to look for one of two items at the same time (Fig. 2). We used a rapid serial visual presentation (RSVP) paradigm, where visual objects are presented in quick succession on a computer screen. In three experiments we investigated if observers can search for (1) more than one shape (Shapes experiment), (2) more than one color (Colors experiment), and (3) one shape and one color at the same time (Combined experiment).

Methods

Participants

Five subjects participated in experiment 1 (3 women, age 19–26). Five new subjects including one of the authors (RH) participated in experiment 2 (4 women, age 18–24). RH also participated in experiment 3, together with seven new subjects (7 women, age 18–34). We discarded the data of three subjects in experiment 3 because their performance was too poor for our analysis (we note that their results were in accordance with our conclusions, see Table 3). All reported normal or corrected vision and gave informed consent. The subjects (except RH) were naive about the purpose of the experiments.

Apparatus and stimuli

The subjects sat in a dimly lit room, 78 cm in front of the stimulus-monitor. The set of stimuli consisted of eight different shapes that we selected from a standardized stimulus-set (Snodgrass & Vanderwart, 1980). The items were relatively dense (with many bright pixels) so that their colors would be easy to discriminate. The colors (red, dark-blue, green, light-blue, yellow, purple, gray, or orange) were equiluminant as determined for each subject. The shapes had a mean width of 2.2° and a mean height of 2.1° and were presented on a black background.
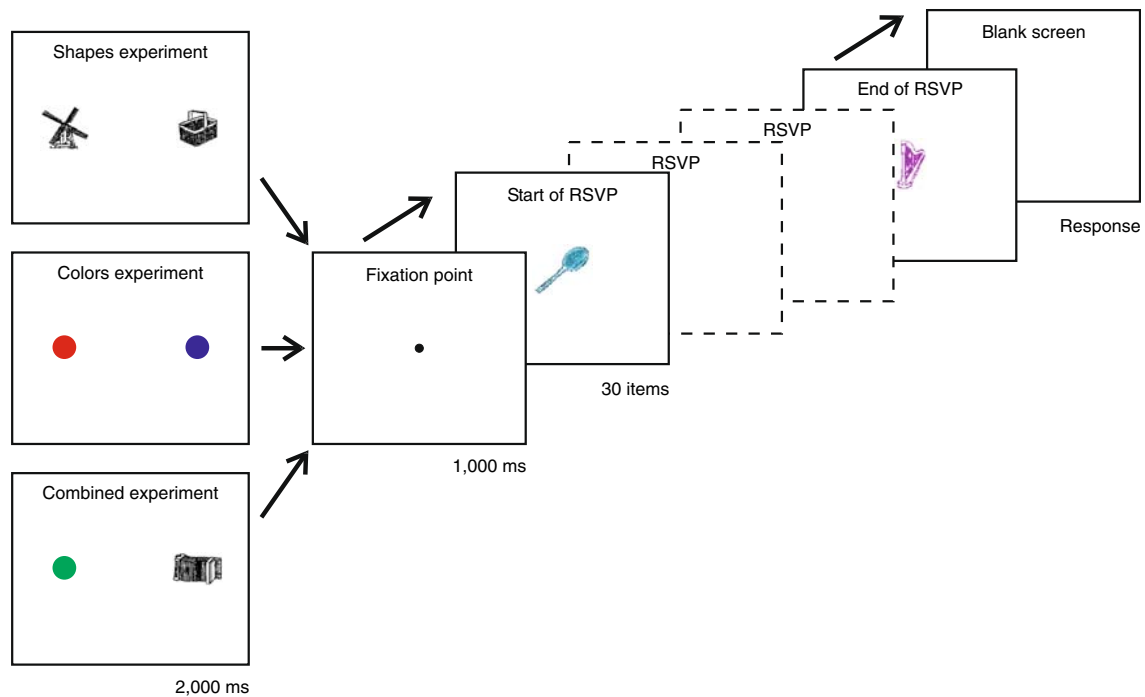
Fig. 2 Sequence of events during a two-target trial. The trial started with a search-target display (not drawn to scale) with either two shapes (Shapes experiment), two colors (Colors experiment) or a color and a shape (Combined experiment) that was presented for 2,000 ms. After an interval of 1,000 ms, an RSVP stream was presented with 30 colored shapes. In 50% of the trials, a single target appeared in the stream, and in the other trials all items were distractors

Procedure

The trial started with a search-target display for 2,000 ms (Fig. 2). On two-target trials two randomly chosen targets were presented, one on the left and one on the right half of the screen. On one-target trials (50% of the trials), the same item was presented on the left and right. The subjects knew that at most one target would ever be present in the stream, and it was their task to indicate whether the stream contained a single target (50% of trials) or none. We did not include trials with more than one target to avoid processing limitations (attentional blinks) that occur if multiple targets appear successively in a stream (Duncan, Ward & Shapiro, 1994; Raymond, Shapiro & Arnell, 1992).

After a fixation point presented for 1,000 ms, a continuous stream of 30 colored shapes was shown in the center of the screen. The shape and color of the distractor items were chosen at random with replacement for every position in the stream with the restriction that a shape or color could never appear twice in a row. The target never appeared at the first three or last three positions of the stream. At the end of the stream the subjects indicated if a target had been included or not by pressing a button. They heard a beep if they made an error.

Every subject started with a baseline condition where the task was to search for a single shape in the Shapes and Combined experiments, and a single color in the Colors experiment. We used a staircase procedure (Wetherill & Levitt, 1965) to determine the presentation rate at which performance was at threshold (84% correct). The resulting mean presentation rate of the items across subjects was 82 ms (60–130 ms) in the Shapes experiment, 106 ms (80–170 ms) in the Colors experiment, and 90 ms (70–120 ms) in the Combined experiment.

Results

Experiment 1

Here observers looked for one or two target shapes in a stream of colored objects (Fig. 2; Shapes experiment). The difference in performance between the two conditions was dramatic; it dropped from an average of 90% on one-target trials to 65% on two-target trials (Fig. 3a, Table 1). This difference was significant in all subjects (chi-square test, χ^2 (1) > 11.00, $P < 0.001$, in all cases). However, a decrease in performance on two-target trials is not necessarily caused by limited target matching capacity. Two concurrent matching processes are expected to give rise to poorer performance than a single process of the same fidelity, because both processes may cause false alarms (see e.g. Verghese, 2001; Wilken & Ma, 2004; see also Greenlee & Thomas, 1993; Magnussen, Greenlee &
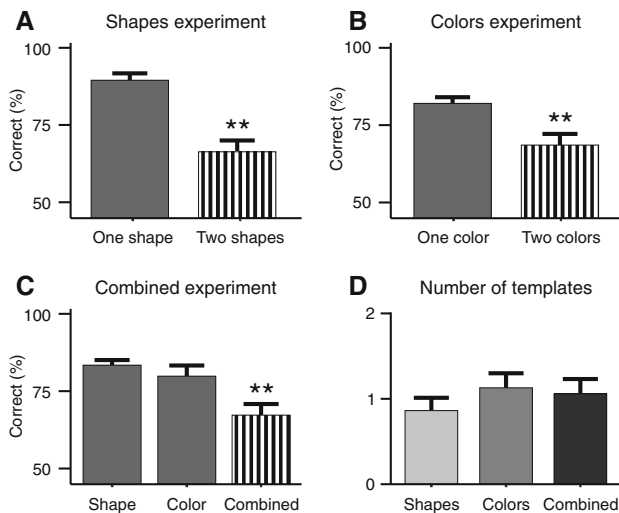
Fig. 3 Performance in the three experiments. **a–c** Percentage of correct responses in the Shapes experiment (**a**), the Colors experiment (**b**), and the Combined experiment (**c**). *Gray bars* show performance for one-target trials, and *striped bars* for two-target trials. *Asterisks* indicate a significantly lower performance on two-target trials than on one-target trials ($P < 0.001$). **d** Estimated average number of active templates in the two-target trials of the Shapes, Colors, and Combined experiment. *Error bars* indicate standard deviation across subjects

Table 1 Mean percentage correct for the five participants of the Shapes experiment for one-target and two-target trials, together with the estimated number of templates

Participant	One target	Two targets	N templates
S1	85.5	56.8	0.52
S2	85.8	68.4	0.96
S3	91.7	63.5	0.75
S4	95.2	61.2	0.85
S5	90.7	73.0	1.33

Thomas 1996). It is possible to correct for this effect because if one matching process has a false alarm rate of f_1, then the joint false alarm rate of two parallel detection processes equals $f_2 = 1 - (1 - f_1)^2$. The equivalent relation for the hit-rate of the combined matching process is $h_2 = 1 - (1 - h_1)(1 - f_1)$, where h_1 is hit-rate of one individual matching process (see appendix A for a derivation of these equations). In addition, we corrected for the possibility that subjects might have a different bias (i.e. the probability to report 'target present') on two-target trials than on one-target trials, by using the logic of the signal detection theory (SDT; Green & Swets, 1966).

We thus derived a 'two-template model', which holds that the subjects can perform two simultaneous matching processes. We assumed that the d' (signal strength in SDT), in two target trials was the same as in one-target trials, while we allowed the response bias, λ, to differ between trial types (see Appendix A for details). The continuous

curves in Fig. 4 shows the predicted relation between hit-rate and false alarm rate as a function of λ, for each participant. It can be seen that the two-template model overestimates the performance of all subjects. We conclude that the decrease in performance can neither be explained by the increase in false alarm rate associated with an additional detection process, nor by a change in the subjects' bias.

We therefore considered a one-template model, which assumes that only a single memory-item can be matched against the visual input at a time. In this model, the subject's performance is equal to that on one-target trials if the RSVP stream happens to contain the item that matches the active template. But if the other target appears in the stream, performance is at chance. The accuracy of all subjects was closer to the prediction of the one-template model (dashed curves in Fig. 4) than to that of the two-template model.

We next estimated the number of active templates in every subject by fitting their performance to a mixture model. If subjects were better than predicted by the one-template model we assumed that they used two templates on a fraction p_2 of the two-target trials and only one template on the other trials, and estimated the average number of templates as $1 + p_2$. If subjects performed worse than predicted by the one-template model, we assumed that zero templates were used on a fraction of the trials (p_0), as could happen, for example, during a switch from one active template to the other one. In that case the average number of templates was estimated as $1 - p_0$.

The average number of templates was 0.9 (Fig. 3d, data of individual subjects are shown in Table 1), a value that did not differ significantly from 1 [t test, $t(4) = 0.93$, $P > 0.4$], but was lower than 2 [$t(4) = 7.80$, $P < 0.01$]. We conclude that effectively only a single shape in working memory acted as search-template at a time. It is unlikely that this limitation is caused by the inability of subjects to store both targets in memory. Alvarez & Cavanagh (2004) measured subjects' capacity to memorize similar items using a change detection procedure and found that it was larger than two. Furthermore, in a previous study we found that subjects were well able to memorize two items of the same stimulus set while comparing one of them to the visual input (Houtkamp & Roelfsema, 2006).

Experiment 2

The shapes that were used as targets in the first experiment were fairly complex. Visual search studies suggest that colors are easier to detect than shapes (e.g. Motter & Belky, 1998), and in the second experiment we explored the possibility that more than one template at a time can support target detection when targets are defined by color.

Fig. 4 Comparison between different models and the subjects' performance in the two target trials of Experiment 1 (Shapes). *Continuous curves* show the predicted relationship between the hit-rate p(hit) and false alarm rate p(false alarm) of a two-template model. *Dashed curves* show predictions of a one-template model. Predicted performance was derived from the sensitivity (d') on one-target trials while the bias (λ) was varied. *Black dots* show the subjects' actual performance on two target trials. *Numbering* of the subjects corresponds to the numbering in Table 1
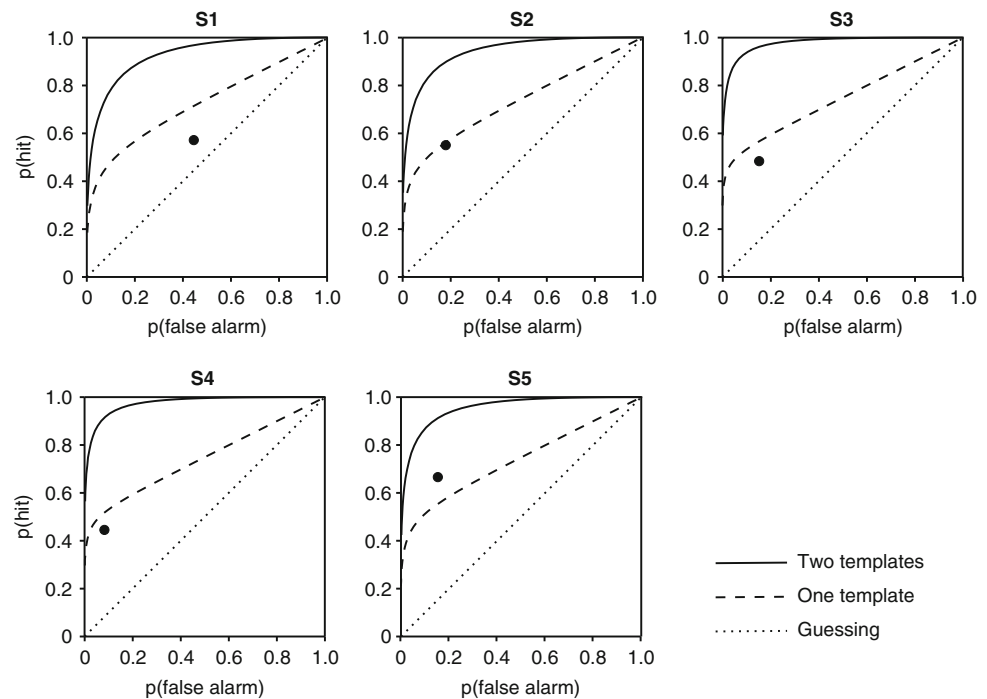
We used the same RSVP stream (Fig. 2; Colors experiment) and asked another group of subjects to look for one or two colors.

Performance decreased from an average of 82% correct on the one-target trials to 69% on two-target trials (Fig. 3b, Table 2). The difference was highly significant for four out of five subjects [χ^2 (1) > 11.23, P < 0.001], while there was a trend in the same direction for the last subject [χ^2 (1) = 2.31, P < 0.07]. The estimated number of templates was 1.1, on average (Fig. 3d). This value was lower than 2 [t test, $t(4) = 5.07$, P < 0.01], but did not differ significantly from 1 [$t(4) = 0.76$, P > 0.4]. Thus, the capacity to match simple colors in working memory with the input is also limited.

Experiment 3

In the first two experiments, targets were defined on the same feature dimension, i.e. both were shapes or both were colors. We next asked whether the observed interference

only occurs if two search-templates are defined in the same feature dimension, or whether there is a more general limitation that even occurs if subjects have to match features from different dimensions. Two templates that are defined in different feature domains, e.g. a color and a shape, might be more compatible with each other and suited to support target detection at the same time (cf. Bichot, Rossi & Desimone 2005; Wolfe, 1994). Therefore, in a third experiment (Fig. 2; Combined experiment) we used the same stimuli, but asked subjects to look for a single shape, a single color, or a color and a shape.

The accuracy of target detection was 83% on one-shape trials, and 79% on one-color trials (for inclusion criteria of subjects see Table 3). It decreased to 67% on two-target trials (Fig. 3c), a difference that was significant in every subject [χ^2 (1) > 6.00, P < 0.01, all subjects]. We adapted our procedure to estimate the number of active templates because the signal strength (d') for color and shape detection may differ (see Appendix B for details) and obtained an average number of active templates of 1.1 (Fig. 3d), which was not significantly different from 1 [t test, $t(4) = 0.33$, P > 0.7] but lower than 2 [$t(4) = 5.47$, P < 0.01). We conclude that subjects are unable to carry out multiple matching processes at the same time even if the features are from different domains.

Discussion

We asked subjects to search for two items at the same time, and found that performance was much poorer than when

Table 2 Percentage of correct responses for one-target and two-target trials and estimated number of templates for the participants in the Colors experiment

Participant	One target	Two targets	N templates
S1	80.6	73.9	1.73
S2	79.3	63.3	0.84
S3	84.7	61.6	0.77
S4	80.7	75.8	1.28
S5	84.3	70.1	0.95

Table 3 Performance of the participants in the Combined experiment for one-shape, one-color, and two-target trials, and the estimated number of templates

Participant	One shape	One color	Two targets	N templates
S1	81.9	84.7	71.7	1.21
S2	82.4	82.4	65.1	1.00
S3	82.1	77.6	67.9	1.35
S4	85.4	80.9	72.4	1.31
S5	82.8	71.1	56.6	0.41
X1[a]	75.0	75.6	69.3	–
X2[b]	84.1	64.1	65.4	–
X3[b]	91.4	60.9	63.0	–

[a] Subject X1 was discarded from further analyses due to a low performance on the one-shape trials (75%), in spite of the fact that performance was approximately 84% in the preceding baseline experiment with only the one-shape task

[b] We expected some variability in the one-color task for the subjects in this experiment, because we adjusted the presentation rate in the baseline experiment according to the subject's performance in the shape detection task. However, two subjects (X2 and X3) performed so poorly on the one-color trials that their data was analyzed separately from the data of the other subjects, who were better than 70% ($d' > 1$) on the one-color trials. The template analysis in subjects X2 and X3 did not converge to a unique solution due to their low performance on the one-color trials ($d' < 1$)

they had to look for a single target. The observed decrease in accuracy was significantly larger than predicted by two parallel matching processes with the same accuracy, but it was compatible with the subjects using only a single search-template at a time.

It is well known that behavioral data usually cannot distinguish between processes that have to be executed in series and processes that are executed in parallel but that share the same, limited resource (Townsend, 1999). Thus, although our data show that the capacity of the matching process is effectively limited to one template (subjects perform as good on two-target trials as they would have if they had used only one template at a time), we cannot exclude that there were in fact two parallel processes with reduced accuracy. Nevertheless, we believe that a process that uses only a single template at a time provides a more parsimonious account of the data. It is remarkable how close the number of estimated templates was to one, in each of the three experiments. Had there been multiple parallel templates sharing the same limited resource, the effective number of templates could have been any value between 1 and 2. Seriality of the matching process is also in line with the results of a seminal study by Sternberg (1966), who investigated the time that subjects require to match a number of characters in memory to a single character that they saw. Subjects' reaction time increased linearly with the number of memory-items, and Sternberg therefore conjectured that subjects perform a serial scan through the items in

memory. Our new method to estimate the number of active templates proves his conjecture: effectively only one item can be matched at a time. This method also allows us to go beyond a previous study (Schneider & Shiffrin, 1977) that demonstrated poorer matching capacity with larger memory set sizes in an RSVP stream, but that did not directly measure the number of active templates.

The limited matching capacity can explain why subjects require more time if they have to search for multiple targets in conventional search tasks (Linnell & Humphreys, 2002; Moore & Osman, 1993; Quinlan & Humphreys, 1987). In this situation subjects will have to switch between search-templates and match one at a time. It also suggests why some studies show only a weak effect of items in working memory on the deployment of attention during visual search (Downing & Dodds, 2004; Houtkamp & Roelfsema, 2006). Apparently there can only be one active template at a time, and the accessory memory-items are in a more passive state with little influence on the deployment of attention.

This idea receives additional support from the previous study by Houtkamp and Roelfsema (2006), where subjects were asked to first search for item A in one display and then for item B in a second display. Subjects were much faster in the second display when A and B were the same than when they were different. In the latter condition subjects had to switch between search-templates, and it apparently takes time to change a passive memory-item into an active search-template (cf. Wolfe et al., 2004).

The present results combined with these earlier studies could be of relevance for working memory theories that distinguish between storage mechanisms and executive processes (e.g. Logie, 1995; Smith & Jonides, 1999). One of the proposed functions of the executive processes is to select relevant sensory input (Smith & Jonides, 1999). The new results, taken together, suggest that the executive processes use only a single memory item at a time (corresponding to the search-template), while working memory can store multiple accessory items in a more passive state.

Our second novel finding is that subjects are unable to simultaneously match features of different categories (a color and a shape). This finding may, at first sight, seem to be inconsistent with recent theories and data about visual search. Specifically, there is neurophysiological data (Bichot et al., 2005) as well as psychophysical data (e.g. Wolfe, Cave, & Franzel, 1989) showing that features from different dimensions can simultaneously guide attention to the target object during visual search. Some of these findings inspired the guided search model (Wolfe, 1994), which holds that multiple features can simultaneously exert a top-down influence on the deployment of attention. Close scrutiny of this model shows, however, that the discrepancy is only apparent. Guided search addresses visual search in

displays with multiple items. For the present discussion it is of importance to consider two of its processing steps. At an early step, the representations of the objects in the visual display compete for selection by attention. The various features of the search-template bias visual selection in parallel, so that the visual object that shares most features with the template has the highest probability to be selected by attention. This step is followed by a second phase where the selected object is matched against the representation of the target object in memory, a matching process that occurs for only one object at a time.

Thus, the search template plays a dual role in guided search: it guides the selection process at the first stage and it is matched against the selected display item at the second. We note that in our experiments the items were presented one by one, so that the first selection step was unnecessary. The items only had to be matched to the template(s) and it is this process that was shown to have a limited capacity. In other words, our findings are not inconsistent with guided search, but rather prove one of its assumptions: the matching phase can occur only for a single object at a time. It will be of considerable interest for future research to elucidate the relationship between the guidance of selective attention and the matching process.

Despite the massively parallel architecture of the visual system where different features are processed in different brain regions (Felleman & Van Essen, 1991), the processing bottleneck observed in the present study is not without precedent. Similar bottlenecks are observed when subjects attempt to detect two targets that are closely separated in time in an RSVP stream (the attentional blink; Duncan et al., 1994; Raymond et al., 1992), or more generally, when subjects try to perform two tasks at the same time (the psychological refractory period; Jolicoeur & Dell'Acqua, 1999; Pashler, 1984). We propose that matching belongs to the set of processes in vision that have a limited capacity.

Appendix A

We estimated the signal strength d' on one-target trials by applying the concepts of the signal detection theory (Green & Swets, 1966):

$$\hat{d}' = Z(h_1) - Z(f_1) \tag{1}$$

h_1 is the proportion of hits, and f_1 is the proportion of false alarms. Z is the inverse of the cumulative standard Gaussian distribution (with mean 0 and variance 1),

$$Z(x) = \Phi^{-1}(x), \quad \text{with} \quad \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \text{Exp}(-0.5z^2) dz \tag{2}$$

The subject's threshold equals λ, so that

$$f_1 = 1 - \Phi(\lambda_1) \tag{3}$$

$$h_1 = 1 - \Phi(\lambda_1 - d'). \tag{4}$$

The following two paragraphs describe how a one and a two template model would predict the subject's performance.

Performance with two templates

In the two-template model the overall false alarm rate, f_2, depends on f_1, the false alarm rate of the individual detection processes, as follows:

$$f_2 = 1 - (1 - f_1)^2 = 1 - \Phi(\lambda_2)^2. \tag{5}$$

On a target present trial, a miss can only occur if the process with the relevant template as well as process with the other, irrelevant template do not reach their thresholds, with probabilities of $1 - h_1$ and $1 - f_1$, respectively. The probability of a miss therefore equals $(1 - h_1)(1 - f_1)$ and we can compute h_2, the hit-rate in the two template trials as follows:

$$h_2 = 1 - (1 - h_1)(1 - f_1) = 1 - \Phi(\lambda_2 - d')\Phi(\lambda_2) \tag{6}$$

where h_1 is the hit-rate of the single detection process. d' was estimated from the one-template trials. Equations 5 and 6 show that f_2 and h_2 depend on one free model parameter, λ_2, the bias on two-target trials. Every value of λ_2 corresponds to a unique combination of f_2 and h_2 (continuous lines in Fig. 4, main text).

Performance with one template

Here d' is equal to that on one-target trials if the target corresponds to the active template, otherwise the subject behaves as on a target-absent trial. It follows from Eq. 3 that the false alarm rate (with the new λ_2) is:

$$f_2 = 1 - \Phi(\lambda_2) \tag{7}$$

The hit-rate h_2 equals h_1 on the 50% of target-present trials where the target is presented that corresponds to the active

template, and f_1 on the other 50% of trials where the wrong template is active.

$$h_2 = 0.5 \cdot (1 - \Phi(\lambda_2)) + 0.5 \cdot (1 - \Phi(\lambda_2 - d')). \qquad (8)$$

Equations 7 and 8 imply that every value of λ_2 is associated with a unique combination of f_2 and h_2 (dashed lines in Fig. 4).

Modeling the data of the subjects

To fit the performance of subjects who performed better than predicted by the one-template model, we assumed that they used two templates on a fraction p_2 of trials, and only a single template on the other trials. By combining Eqs. 5–8, the predictions of the mixture model are as follows:

$$f_2^{\text{Mixture}} = p_2\left(1 - \Phi(\lambda_2)^2\right) + (1 - p_2)(1 - \Phi(\lambda_2)) \qquad (9)$$

$$\begin{aligned} h_2^{\text{Mixture}} = {}& p_2(1 - \Phi(\lambda_2 - d')\Phi(\lambda_2)) + (1 - p_2) \\ & \times (0.5 \cdot (1 - \Phi(\lambda_2)) + 0.5 \cdot (1 - \Phi(\lambda_2 - d'))). \end{aligned}$$
$$(10)$$

These two equations have two unknown parameters, λ_2 and p_2, and they can therefore be solved numerically. The average number of templates used by the subject is estimated as

$$N\_\text{Templates} = 1 + p_2. \qquad (11)$$

We assumed that the subjects performing worse than predicted by the one-template model had no active template on a fraction p_0 of trials and one template on the other trials. The model assumed that the subject pressed the 'yes' button with the same probability, $p(\text{Yes}|\text{No Template})$, as on one-template trials (this is not a critical assumption). Thus:

$$p(\text{Yes}|\text{No Template}) = 0.5h_2^{\text{One Template}} + 0.5f_2^{\text{One Template}} \qquad (12)$$

$f_2^{\text{One Template}}$ and $h_2^{\text{One Template}}$ were defined in Eqs. 7, 8; note that they only depend on λ_2. f_2 and h_2 can be computed by combining Eq. 12 with Eqs. 7 and 8:

$$f_2^{\text{Mixture2}} = (1 - p_0)(1 - \Phi(\lambda_2)) + p_0 p(\text{Yes}|\text{No Template}) \qquad (13)$$

$$\begin{aligned} h_2^{\text{Mixture2}} = {}& (1 - p_0)(0.5 \cdot (1 - \Phi(\lambda_2)) \\ & + 0.5 \cdot (1 - \Phi(\lambda_2 - d'))) \\ & + p_0 p(\text{Yes}|\text{No Template}). \end{aligned} \qquad (14)$$

Equations 12–14 have three unknown parameters, λ, p_0, and $p(\text{Yes}|\text{No Template})$ and can be solved numerically.

For these subjects, $(1 - p_0)$ was used as an estimate of the average number of active templates (see Tables 1, 2 for the estimated number of templates).

Appendix B

The combined experiment requires a more complex model, because the matching processes for color and shape may have different sensitivities d'_S and d'_C and biases λ_S and λ_C.

Performance with two templates

The combined false alarm rate f_2 can be computed on the basis of the individual rates f_S and f_C:

$$f_2 = 1 - (1 - f_S)(1 - f_C). \qquad (15)$$

We compute the hit-rate on trials with the shape- and color-target as:

$$h_{2S} = 1 - (1 - h_S)(1 - f_C) \qquad (16)$$

$$h_{2C} = 1 - (1 - h_C)(1 - f_S). \qquad (17)$$

Two parameters of the model, λ_C and λ_S, determine three observables, f_2, h_{2S} and h_{2C}. The model can be tested because two of these observables predict the third.

Performance with one template

According to the one-template model, the shape template is selected with probability p_S, and the color template with probability $p_C = 1 - p_S$. The subject's d' is equal to that on one-target trials if the target matches the active template. Otherwise, the subject behaves as if the trial is a target-absent trial. The false alarm rate is as follows:

$$f_2 = p_S(1 - \Phi(\lambda_S)) + (1 - p_S)(1 - \Phi(\lambda_C)) \qquad (18)$$

and the hit-rates equal

$$h_{2S} = (1 - p_S) \cdot (1 - \Phi(\lambda_C)) + p_S \cdot (1 - \Phi(\lambda_S - d'_S)) \qquad (19)$$

$$h_{2C} = p_S \cdot (1 - \Phi(\lambda_S)) + (1 - p_S) \cdot (1 - \Phi(\lambda_C - d'_C)). \qquad (20)$$

Thus in this model, f_2, h_{2S} and h_{2C} are jointly determined by λ_S, λ_C and p_S. d'_S and d'_C together constrain the combinations of f_2, h_{2S} and h_{2C} that can be attained by the model. We note that the subject's accuracy comes close to chance level if one λ is set to a high value and the other to a low value, while it can be shown (using the method of Lagrange multipliers) that the model's performance is optimal when the following relation holds:

$$\phi(\lambda_S)\Big\{\Big[\Phi(\lambda_C)-\Phi(\lambda_S-d'_S)\Big]$$
$$\times[\phi(\lambda_C)-\phi(\lambda_C-d'_C)]-\Big[\Phi(\lambda_C)-\Phi(\lambda_C-d'_C)\Big]\phi(\lambda_C)\Big\}$$
$$=\phi(\lambda_S-d'_S)\Big\{\Big[\Phi(\lambda_C)-\Phi(\lambda_S)\Big]\Big[\phi(\lambda_C)-\phi(\lambda_C-d'_C)\Big]$$
$$-\Big[\Phi(\lambda_C)-\Phi(\lambda_C-d'_C)\Big]\phi(\lambda_C)\Big\}. \tag{21}$$

Here ϕ denotes the normal probability density and Φ the cumulative normal probability density. Analysis of this equation shows that performance is optimal if λ_S is an (almost) linear function of λ_C, and that λ_S should equal λ_C if d'_S equals d'_C. Psychophysical evidence corroborates this relation: subjects select similar decision criteria if they concurrently perform multiple detection processes (Gorea & Sagi, 2000).

Modeling the data of the subjects

For subjects whose performance fell between the predictions of the one- and two-template model, we assumed that they used two templates on a fraction of trials p_2, and only one template on the other trials (with an optimal combination of λs). The following three equations can be derived from Eqs. 15–17, and 18–20:

$$f_2^{\text{Mixture}}=(1-p_2)\{(1-p_S)(1-\Phi(\lambda_C)+p_S(1-\Phi(\lambda_S)\}$$
$$+p_2\{1-\Phi(\lambda_S)\Phi(\lambda_C)\}$$
$$\tag{22}$$

$$h_{2S}^{\text{Mixture}}=(1-p_2)\Big\{(1-p_S)(1-\Phi(\lambda_C)+p_S(1-\Phi(\lambda_S-d'_S)\Big\}$$
$$+p_2\Big\{1-\Phi(\lambda_S-d'_S)\Phi(\lambda_C)\Big\}$$
$$\tag{23}$$

$$h_{2C}^{\text{Mixture}}=(1-p_2)\Big\{(1-p_S)(1-\Phi(\lambda_C-d'_C)+p_S(1-\Phi(\lambda_S)\Big\}$$
$$+p_2\Big\{1-\Phi(\lambda_S)\Phi(\lambda_C-d'_C)\Big\}.$$
$$\tag{24}$$

Together with Eq. 21 there are four equations with four unknown variables, λ_C, λ_S, p_S, and p_2 that can be solved numerically.

For the other subjects we assumed no active template on a fraction of trials p_0, and one template on the other trials. We assumed that they pressed the 'yes' button with the same probability as in the one-template trials. Thus:

$$p(\text{Yes}|\text{No Template})=0.25h_{2S}^{\text{One Template}}+0.25h_{2C}^{\text{One Template}}$$
$$+0.5f_2^{\text{One Template}}$$
$$\tag{25}$$

where $f_2^{\text{One Template}}$, $h_{2S}^{\text{One Template}}$, and $h_{2C}^{\text{One Template}}$ are as defined in Eqs. 18–20 and depend on λ_S, λ_C and p_S. The

hit-rates and the false alarm rate of the full model can be derived by combining Eq. 25 with Eqs. 18–20:

$$f_2^{\text{Mixture2}}=(1-p_0)\{(1-p_S)(1-\Phi(\lambda_C))+p_S(1-\Phi(\lambda_S)\}$$
$$+p_0p(\text{Yes}|\text{No Template})$$
$$\tag{26}$$

$$h_{2S}^{\text{Mixture2}}=(1-p_0)\Big\{(1-p_S)(1-\Phi(\lambda_C))$$
$$+p_S(1-\Phi(\lambda_S-d'_S))\Big\}+p_0p(\text{Yes}|\text{No Template})$$
$$\tag{27}$$

$$h_{2C}^{\text{Mixture2}}=(1-p_0)\Big\{(1-p_S)(1-\Phi(\lambda_C-d'_C))$$
$$+p_S(1-\Phi(\lambda_S)\Big\}+p_0p(\text{Yes}|\text{No Template}).$$
$$\tag{28}$$

Equations 25–28 together with the optimality constraint (Eq. 21) have five unknown variables, λ_C, λ_S, p_S, $p(\text{Yes}|\text{No Template})$ and p_0, and can be solved numerically (see Table 3 for the estimated number of templates).

References

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, *15*, 106–111.

Bichot, N. P., Rossi, A. F., & Desimone, R. (2005). Parallel and serial neural mechanisms for visual search in macaque area V4. *Science*, *308*, 529–534.

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

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

Downing, P. E., & Dodds, C. M. (2004). Competition in visual working memory for control of search. *Visual Cognition*, *11*, 689–703.

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

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, *1*, 1–47.

Gorea, A., & Sagi, D. (2000). Failure to handle more than one internal representation in visual detection tasks. *Proceedings of the National Academy of Sciences of the United States of America*, *97*, 12380–12384.

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

Greenlee, M. W., & Thomas, J. P. (1993). Simultaneous discrimination of the spatial frequency and contrast of periodic stimuli. *Journal of the Optical Society of America*, *10*, 395–404.

Houtkamp, R., & Roelfsema, P. R. (2006). The effect of items in working memory on the deployment of attention and the eyes during visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 423–442.

Jolicoeur, P., & Dell'Acqua, R. (1999). Attentional and structural constraints on visual encoding. *Psychological Research*, *62*, 154–164.

Linnell, K. L., & Humphreys, G. W. (2002). Visual search within and across dimensions: A case for within-dimension grouping. *British Journal of Psychology*, *93*, 115–135.

Logie, R. H. (1995). *Visuo-spatial working memory*. Hove, East Sussex: Lawrence Erlbaum.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279–281.

Magnussen, S., Greenlee, M. W., & Thomas, J. P. (1996). Parallel processing in visual short-term memory. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 202–212.

Moore, C. M., & Osman, A. M. (1993). Looking for two targets at the same time: One search or two? *Perception & Psychophysics*, *53*, 381–390.

Motter, B. C., & Belky, E. J. (1998). The guidance of eye movements during active visual search. *Vision Research*, *38*, 1805–1815.

Oh, S. H., & Kim, M. S. (2003). The guidance effect of working memory load on visual search. *Journal of Vision*, *3*, 629a.

Olivers, C. N. L., Meijer, F., & Theeuwes, J. (2006). Feature-based memory-driven attentional capture: Visual working memory content affects visual attention. *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 1243–1265.

Olivers, C. N. L. (2008). Interactions between visual working memory and visual attention. *Frontiers in Bioscience*, *13*, 1182–1191.

Pashler, H. (1984). Processing stages in overlapping tasks: Evidence for a central bottleneck. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 358–377.

Quinlan, P. T., & Humphreys, G. W. (1987). Visual search for targets defined by combinations of color, shape, and size: An examination of the task constraints on feature and conjunction searches. *Perception & Psychophysics*, *41*, 455–472.

Raymond, J. E., Shapiro, K., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 849–860.

Schneider, W. X., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, *84*, 1–66.

Smith, E. E., & Jonides, J. (1999). Storage and executive processes in the frontal lobes. *Science*, *283*, 1657–1661.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 174–215.

Soto, D., Heinke, D., Humphreys, G. W., & Blanco, M. J. (2005). Early, involuntary top-down guidance of attention from working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *31*, 248–261.

Sternberg, S. (1966). High-speed scanning in human memory. *Science*, *153*, 652–654.

Townsend, J. T. (1999). Serial vs. parallel processing: Sometimes they look like tweedledum and tweedledee but they can (and should) be distinguished. *Psychological Science*, *1*, 46–54.

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

Wetherill, G. B., & Levitt, H. (1965). Sequential estimation of points on a psychometric function. *The British Journal of Mathematical and Statistical Psychology*, *18*, 1–10.

Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, *4*, 1120–1135.

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

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 419–433.

Wolfe, J. M., Horowitz, T. S., Kenner, N., Hyle, M., & Vasan, N. (2004). How fast can you change your mind? The speed of top-down guidance in visual search. *Vision Research*, *44*, 1411–1426.

Woodman, G. F., & Luck, S. J. (2007). Do the contents of visual working memory automatically influence attentional selection during visual search? *Journal of Experimental Psychology: Human Perception and Performance*, *33*, 363–377.