

# Temporal buffering and visual capacity: The time course of object formation underlies capacity limits in visual cognition

Andreas Wutz · David Melcher

Published online: 9 April 2013

© The Author(s) 2013. This article is published with open access at Springerlink.com

**Abstract** Capacity limits are a hallmark of visual cognition. The upper boundary of our ability to individuate and remember objects is well known but—despite its central role in visual information processing—not well understood. Here, we investigated the role of temporal limits in the perceptual processes of forming “object files.” Specifically, we examined the two fundamental mechanisms of object file formation—individuation and identification—by selectively interfering with visual processing by using forward and backward masking with variable stimulus onset asynchronies. While target detection was almost unaffected by these two types of masking, they showed distinct effects on the two different stages of object formation. Forward “integration” masking selectively impaired object individuation, whereas backward “interruption” masking only affected identification and the consolidation of information into visual working memory. We therefore conclude that the inherent temporal dynamics of visual information processing are an essential component in creating the capacity limits in object individuation and visual working memory.

**Keywords** Object individuation · Object identification · Sensory memory · Visual working memory · Capacity · Visual masking

One of the fundamental goals of perception is to enable us to interact with objects in the environment. According to Wundt, the interaction of an observer with the external environment (the “psychophysical process”) can be subdivided into three temporally successive and distinct

stages (Wundt, 1899, 1900). The first stage (“perception”) describes the entrance of an object into the field of vision, allowing it to be detected. In a subsequent stage, termed “apperception,” the perceived object occupies the focus of the observer’s attention. Finally, the observer develops the volition to react to the object, either cognitively, by storing it into memory, or behaviorally, with a grasping or a saccadic eye movement.

Wundt’s description emphasizes how object recognition involves a temporal succession of distinct processing stages - from an unlimited in capacity, but fragile, purely bottom-up and in parallel computed sensory representation (iconic memory: Neisser, 1967; Sperling, 1960, 1963) to a capacity limited, durable and cognitively structured visual store (visual short-term memory: Phillips & Baddeley, 1971; Sperling, 1960, 1963) leading to an action that results in an isomorphic one-to-one relation between observer and object.

As is shown in Fig. 1A, Wundt’s stage of apperception can be further subdivided into two processing mechanisms: object individuation and object identification (Xu & Chun, 2009). Individuation involves selecting features from a crowded scene, binding them into a unitary representation, and individuating this spatiotemporal unit from other objects in the image (Kahneman, Treisman, & Gibbs, 1992; Pylyshyn, 1989; Treisman & Gelade, 1980; Xu & Chun, 2009). Object representations at this stage are suggested to be coarse and contain only minimal feature information (Xu & Chun, 2009). Some of these “object files” (Kahneman et al., 1992) are elaborated subsequently during object identification. It is at this stage that identity information becomes available to the observer, and the content of the object files can be consolidated into durable and reportable representations in visual working memory. The number of objects available at this stage is variable, depending on the object complexity, task demands, and representation resolution (Alvarez & Cavanagh, 2004; Xu & Chun, 2009). As

A. Wutz (✉) · D. Melcher  
CIMeC—Center for Mind & Brain Sciences, University of Trento,  
Palazzo Fedrigotti, corso Bettini 31,  
38068 Rovereto, TN, Italy  
e-mail: andreas.wutz-1@unitn.it

individuation precedes identification, the capacity of the latter has its upper bound at the limit of the former (Dempere-Marco, Melcher, & Deco, 2012; Melcher & Piazza, 2011; Piazza, Fumarola, Chinello, & Melcher, 2011).

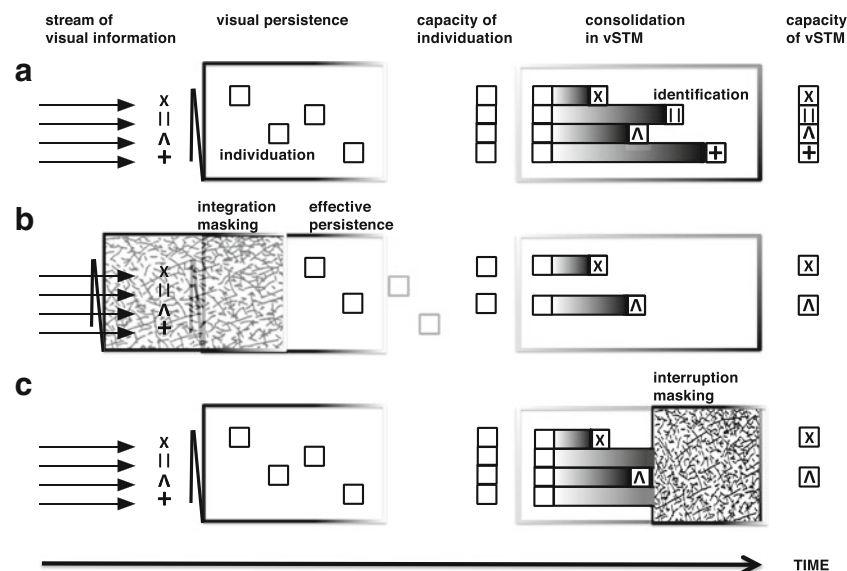
The goal of the present report is to investigate whether capacity limitations in object processing can be traced to temporal constraints on the distinct object-processing stages. We therefore embed the ongoing debate about the roots of capacity limits in vision (reflected in the “subitizing” phenomenon (Jevons, 1871; Kaufman et al. 1949)) and visual working memory (Cowan, 2000; Luck & Vogel, 1997) into the already well-established body of work about the temporal dynamics of the visual system (Loftus, Duncan, & Gehrig, 1992; Sperling, 1960, 1963; Wundt, 1899).

Specifically, we used two types of masking—integration and interruption masking—in order to influence either the individuation or the identification stage of object file formation. *Visual masking* refers to the reduction of the visibility of one stimulus, called the *target*, by another stimulus shown before and/or after it, called the *mask* (Breitmeyer & Ögmen, 2006; Enns & Di Lollo, 2000). This process is usually explained in terms of a two-factor theory, yielding integration and interruption masking (Scheerer, 1973; Scheerer & Bongartz, 1973). *Integration masking* occurs when the target and the mask information are combined, as a consequence of the imprecise temporal resolution of the visual system. Integration masking can occur with either forward or backward masking for short stimulus onset

asynchrony (SOA) values (up to around 100 ms between the target and the mask). In contrast, *interruption masking* affects the higher-level mechanisms that are engaged in object recognition, and it yields a J-shaped masking function, as it can only occur for masks appearing temporally after the target display (Breitmeyer & Ögmen, 2006; Enns & Di Lollo, 2000). The effect of this kind of masking is thought to reflect a disruption of processing *after* perceptual analysis is already completed, but *before* the representation has been consolidated into visual working memory (Vogel, Woodman, & Luck, 2006).

Our hypothesis is that integration masking should selectively affect the individuation stage by reducing the effective persistence of the target items (Fig. 1B). Integration masking is very effectively implemented with a specific forward-masking technique that makes it possible to quantitatively change the duration of visual persistence (and thus of iconic memory access), as well as the degree of temporal integration, by varying the onset asynchrony between the first and second displays (Di Lollo, 1980). Also, in the case of backward masking with a very short SOA, we would expect integration masking to occur and to limit the effective visual persistence of the target, and thus the individuation processes.

In contrast, we predict that interruption masking should selectively affect the identification of items after individuation has largely finished, since the consolidation of targets into visual short-term memory (vSTM) would be interrupted (Fig. 1C). Interruption masking should only occur for



**Fig. 1** Illustration of how the temporal limits of visual object processing can result in capacity limits for individuation and identification. (A) Under normal viewing conditions, the stream of visual information is individuated during the period of visual persistence of the sampled sensory image. Items that are individuated are potential “object files” that can then be identified and consolidated into visual short-term

memory (vSTM). (B) Integration masking via forward masking reduces the effective persistence of the target items, leading to a reduction in capacity for individuation and, consequently, also for identification. (C) Interruption (backward) masking does not influence the initial individuation of items, but instead disrupts the identification and consolidation of items into vSTM

backward masking with longer SOAs (greater than around 100 ms). We would therefore expect to see a specific influence of such backward masking on visual memory, but not on individuation.

We investigated the two stages of object file formation (individuation and identification/consolidation) using the two forms of contour masking (integration and interruption) in a fully counterbalanced two-by-two design. In order to watch the temporal unfolding of object file formation, we employed forward- and backward-masking techniques, using a variety of SOAs, in two tasks: enumeration and change detection. Enumeration served as an operationalization of object individuation, whereas change detection served as the main paradigm for studying visual working memory.

If capacity limits in vision and visual working memory can be explained by temporal constraints on the formation of object files, we would expect that techniques that limit processing time at specific temporal stages of the visual analysis would selectively inhibit the successive mechanisms operating upon the sensory input at these stages. In other words, integration masking should selectively impair object individuation, whereas interruption masking should only affect object identification and the consolidation of object information into visual working memory. This design allowed us to test the role of temporal dynamics in the individuation and identification of objects.

We also included a control condition to measure the effects of the forward- and backward-masking paradigms on a simple detection task. This control condition was necessary in order to ensure that reduced performance from masking did not simply reflect the fact that the targets were effectively invisible, but revealed limits on the visual computations *within* the sensory image aimed to arrive at a structured, object-like representation. This control condition allowed us to study the unfolding of object representations, from simple detection of the presence of a stimulus, to the individuation of a specific number of target items, and eventually to the recognition of object file content.

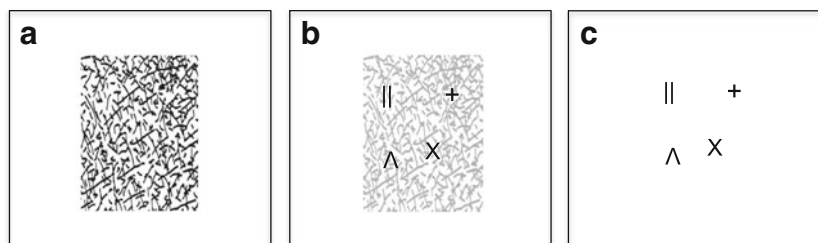
## Method

### Subjects

A group of 16 subjects (11 female, five male; mean age  $M = 22.9$  years,  $SD = 4.2$  years) completed a series of four conditions in the main experiment on object file formation. A different group of ten subjects (six female, four male; mean age  $M = 22.7$  years,  $SD = 3.7$  years) participated in the control condition measuring target detection. All of the subjects provided informed consent, as approved by the institutional ethics committee. Subjects took part in exchange for course credit or a small payment and had normal or corrected-to-normal vision.

### Stimuli and apparatus

The experiment was run on an HP Intel Quad core computer using MATLAB 7.9 (MathWorks, Natick, MA) and the Psychophysics Toolbox, Version 3 (Brainard, 1997; Pelli, 1997). Participants were seated in a dimly lit room, approximately 45 cm from a 19-in. Mitsubishi monitor (1,600 × 1,200 resolution) running at 85 Hz. On each trial, a different pattern of 400 randomly oriented, partially crossing black lines (mean line length =  $1^\circ$  visual angle, mean line width =  $0.1^\circ$ , mean size of whole pattern =  $13.4^\circ$ ) was presented, centered on a white background (Fig. 2A). In the forward-masking conditions, this pattern remained on the screen and then, after a variable onset delay, a variable number of items (up to six) appeared that were linearly superposed upon the random line pattern by use of the image-processing technique “alpha blending” (Fig. 2B). The physical properties of both the mask and target elements—that is, their contrast, mean line length, and mean line width—were equated. Furthermore, the “alpha-blending” procedure was used to edit the transparency/opacity values of the visual stimuli, assuring a mathematically correct superimposition of local element contrast, without creating any discontinuities in contrast that would have been a cue to finding the target. The mask and target elements differed only in their temporal



**Fig. 2** Illustration of the mask and target stimuli used in the experiment. (A) Display with an example random-line pattern used for all of the masks in the experiment. (B) Example of the two-line drawings used as targets, superimposed on the random-line pattern, which is

shown here at 60 % transparency for illustrative purposes. In the experiment, the random-line pattern was always shown at full contrast, as it is in panel A. (C) Example of the two-line drawings used as targets upon a blank white screen, as in the backward-masking conditions

onsets, in order to exclusively vary the amount of integration masking. Thus, this method combined both forward and simultaneous masking. In the backward-masking conditions, the same random-line pattern was presented at a variable interstimulus interval with respect to target offset (Fig. 2A). The same set of 12 possible two-line drawings (i.e., two crossed or parallel lines) was used for the items in all four experimental manipulations, and also in the control conditions (see Figs. 2B and C). All items were colored black, were  $0.9^\circ$  of visual angle in size, and were placed randomly at one of 16 possible locations within an invisible, central rectangle  $5.4^\circ$  of visual angle in eccentricity, with a minimum buffer of  $0.6^\circ$  between the locations.

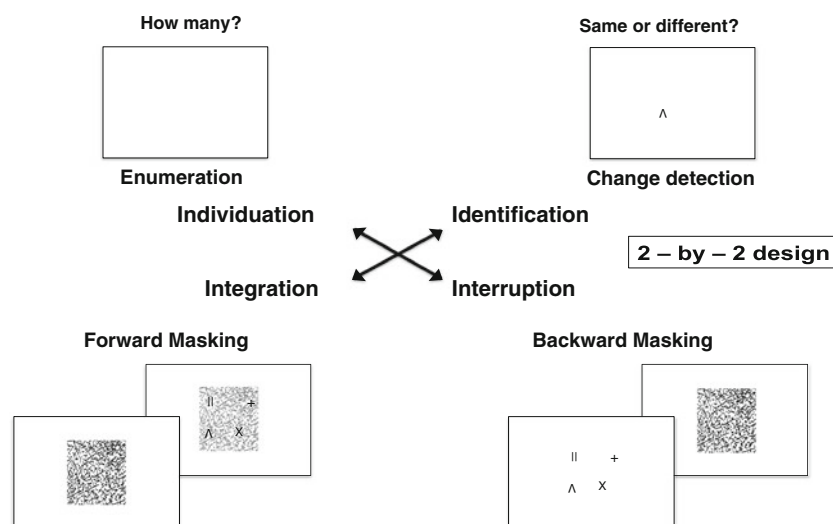
### Procedure

Each subject completed the four experimental manipulations in two sessions consisting of two conditions each and lasting approximately 1.5 h apiece. The serial order of the four different experimental manipulations (masking technique [forward vs. backward] crossed with task [enumeration vs. change detection]) was fully balanced across the observers, in a Latin square design (Fig. 3). Groups of four subjects completed one of the four counterbalanced sequences within the Latin square. Prior to the experiment, the full set of possible target items was presented to the subjects on the screen for an unrestricted viewing time. All subjects received verbal and written instructions about each task and completed 20 practice trials for each condition. In all four conditions, each trial began with a central fixation dot

(black,  $0.3^\circ$ ) on a white background for 500 ms, followed by a blank white screen for another 500 ms. Then, the order of events in the trial depended on the masking technique and task, as is explained below. The subjects' responses on the keyboard initiated the next trial.

*Forward versus backward masking* In the case of forward masking, the random-line pattern was presented for one of four durations, in order to control the SOA between the onset of the mask and the item(s). Four different SOAs were used: 24, 47, 200, or 494 ms. The target display with the items to be enumerated or memorized was superposed upon the masking pattern and was always presented for the same brief duration of 71 ms (Fig. 2B). The target display was immediately followed by a white screen (Fig. 4A). Using this procedure, we achieved an optimal temporal resolution of the visual mechanisms operating within the first tens of milliseconds around target exposure, during which integration masking mostly occurs, as very short SOAs can be used. This simultaneous mask made it possible to fractionate the time course of visible persistence of the target items. It is important to note that rather than merely reducing item visibility, the combination of forward and simultaneous masking specifically affected the rate at which objects were individuated: At short SOAs, only one object could be individuated, while with increasing SOAs, object capacity increased in steps (Wutz, Caramazza, & Melcher, 2012).

On backward-masking trials, the target items were shown first for 71 ms upon a white background, followed by the random-line pattern after a variable SOA. Unlike in the

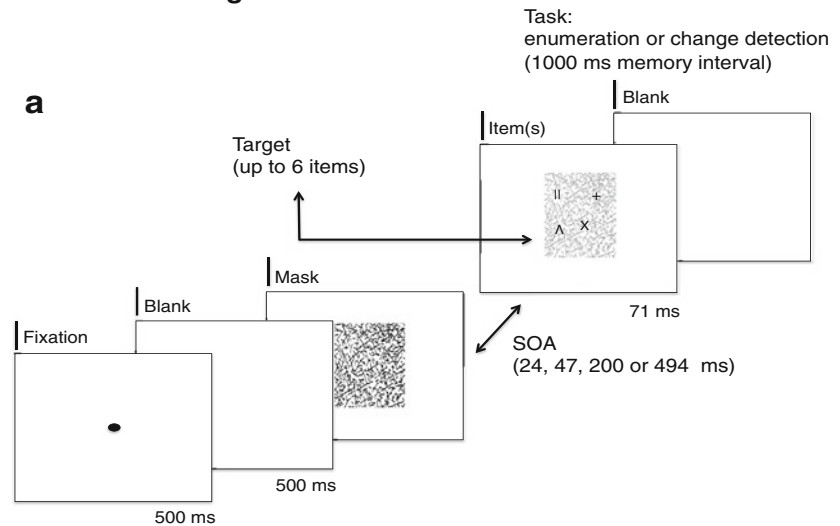


**Fig. 3** Schematic depiction of the  $2 \times 2$  design employed in the experiment. Each of the four conditions (forward masking with enumeration, forward masking with change detection, backward masking with enumeration, and backward masking with change detection) was administered to the subjects using a fully counterbalanced Latin square design. Enumeration served as an operationalization of object

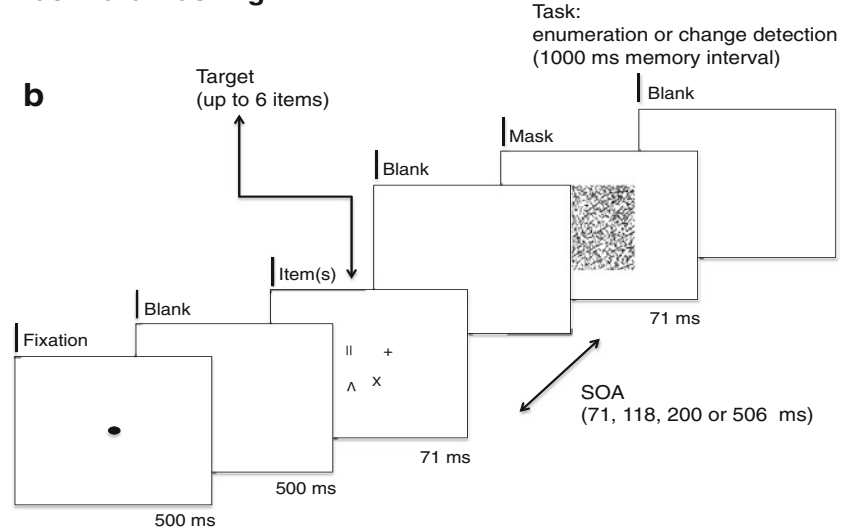
individuation, while change detection measured object identification. The technique of forward masking is considered to favor integration masking (Di Lollo, 1980), and backward masking with longer SOAs has an interrupting influence on visual performance (Scheerer, 1973; Scheerer & Bongartz, 1973)

**Fig. 4** A single trial apiece in the forward-masking (panel A) and backward-masking (panel B) conditions. (A) Illustration of one trial in the forward-masking conditions. Throughout the trials, the two independent factors Target Set Size (1, 2, 3, 4, or 6, for enumeration; 2, 4, or 6, for change detection) and Mask–Target SOA (24, 47, 200, or 494 ms) were varied. The targets superimposed on the masking pattern (here shown 60 % transparent for illustrative reasons) were always presented for 71 ms, followed by a blank screen until the subject’s response in the enumeration condition. During change detection, a memory interval of 1,000 ms followed the target display, followed by a probe item for 71 ms. (B) Illustration of one trial in the backward-masking conditions. Throughout the trials, the two independent factors Target Set Size (1, 2, 3, 4, or 6, for enumeration; 2, 4, or 6, for change detection) and Mask–Target SOA (71, 118, 200, or 506 ms) were varied. The targets were always presented for 71 ms, followed (in the case of SOAs bigger than 71 ms) by a blank screen and a mask for 71 ms. In the enumeration condition, a blank screen followed until the subject’s response. During change detection, a memory interval of 1,000 ms followed the target display, followed by a probe item for 71 ms

### Forward Masking



### Backward Masking



forward-masking technique, the target and masking displays were not presented simultaneously (Fig. 2C). Four different SOAs were used: 71 ms (i.e., immediately after target offset), 118, 200, or 506 ms. Any delay period between target offset and mask onset was filled by the presentation of a blank white screen. The mask was always shown for 71 ms and immediately followed by a white screen (Fig. 4B). The 71-ms SOA mask condition was included so as to fit within the temporal limits of integration masking, while the longer SOAs were expected to result in interruption masking.

*Enumeration versus change detection* Both masking techniques were used in a crossed design with two different task demands: enumeration or change detection within the item display. In the case of enumeration, the subjects had to

indicate the number of perceived items by pressing the corresponding number on a keyboard immediately after target or mask offset. Whereas one to four or six items were actually shown (there were never five targets), the subjects were instructed to respond within the full range between one and six items. We did not inform subjects that none of the trials included five target items, in order to avoid a guessing strategy in which subjects would always respond “six” when the number of items exceeded their subitizing range. The enumeration condition consisted of six blocks of 60 trials, with each of the 20 possible combinations of SOA and target numerosity being shown three times per block in random order.

On change detection trials, a probe was presented, after a blank delay of 1 s, for 71 ms in one of the locations that had



previously been occupied by a target item. This memory interval of 1 s was always held constant, regardless of the temporal position of the mask and the item display. The identity of the probe matched the corresponding item in the target set on 50 % of the trials. Participants responded by pressing a key corresponding to the probe identity being the same or different. Within one block, every combination of the three factors—SOA, Set Size, and Probe Identity—was shown three times and in random order. The conditions involving change detection comprised five blocks of 72 trials.

**Target detection** In order to clearly disentangle the effects of masking on the formation on object files from a more generic effect on target display visibility, we ran a control condition requiring subjects simply to detect the target display. These subjects reported whether or not at least one target had been presented on each trial. Each of the ten subjects was run in this control task under both forward- and backward-masking conditions in a single session. The order of the masking types was balanced across subjects. All of the subjects received verbal and written instructions about the task and completed one practice block for each condition. The trial sequence and the masking procedures used in this control condition were identical to those described above, except for the following changes: Only two SOAs were used—the shortest and longest ones described in the experimental procedures above. This meant that, for forward masking, the SOAs were 24 and 494 ms, while for backward masking, we used SOAs of 71 ms (immediately after target offset) and 506 ms. Target displays were presented on 50 % of the trials. Within these target-present trials, the display consisted with equal probability of either one or four targets presented for 71 ms. On the other half of the trials (target-absent trials), the target display was replaced either with an instance of the masking pattern (on forward-masking trials) or with a white screen (on backward-masking trials) for an equal duration (71 ms). The subjects were instructed to press a previously specified key indicating the presence or absence of a target display, irrespective of the number of targets, after mask or target offset, respectively. Within one block, every combination of the two factors—SOA and Target Presence/Absence—was shown 16 times and in random order. For both forward and backward masking, three blocks of 64 trials each were run. The whole session lasted approximately 45 min.

#### Data analysis

For all experimental conditions, the proportions of correct trials were fed into a two-way within-subjects analysis of variance (ANOVA) with the factors Set Size (1–4 and 6, for enumeration; 2, 4, and 6, for change detection) and SOA. In

the event that the residuals of one variable within one condition did not follow a normal distribution, as indexed by a Kolmogorov–Smirnov test, the analysis for this condition was repeated using a Friedman test. As the main results did not differ between the parametric and nonparametric procedures, only the ANOVA results are reported. If sphericity for a given factor was not tenable, the reported  $F$  ratios have been adjusted with a Greenhouse–Geisser correction. The alpha level for post-hoc planned comparisons has been corrected with a Bonferroni procedure. For better comparability of the results between the different conditions involving object file formation, the proportions of correct trials with set size 4 have been translated into corresponding capacity estimates for each SOA. This calculation was based on the performance measures for the four-item displays in accordance with previous reports (e.g., Vogel et al., 2006), since visual object capacity is likely to converge toward asymptote for this set size (Cowan, 2000). The computation of the capacity estimates takes into account the different guessing rates within the different response measures used (enumeration and change detection). For change detection, capacity  $K$  has been calculated using the following formula:

$$K = (H + CR - 1) \times N,$$

where  $K$  indicates capacity,  $H$  hit rate,  $CR$  correct rejection rate, and  $N$  the number of items in the display (Cowan, 2000).

For enumeration, a guessing correction for a six-alternative forced choice procedure was applied on the raw proportions of correct trials (Klein, 2001). Capacity estimates were then derived by multiplying these values by the number of items in the display, as is explicit in the following formula:

$$K = [(P_{\text{cor}} - 1/M) / (1 - 1/M)] \times N,$$

with  $K$  being capacity,  $P_{\text{cor}}$  the proportion of correct trials,  $M$  the number of alternatives (here, six), and  $N$  the number of items in the display.

## Results

### Visual masking and object file formation

In all four conditions of the main experiment, we found main effects of both set size and SOA on the proportions of correct responses (Table 1). These main effects confirm the evident trend, in Fig. 5, of improved performance for longer SOAs and smaller set sizes. Significant interactions also emerged in three of the four conditions (Table 1). The ordinal order of the main effects, however, was preserved despite these interactions (Fig. 5).

**Table 1** Results of the two-way within-subjects ANOVA of the four crossed conditions in the main experiment: Forward masking with enumeration, forward masking with change detection, backward masking with enumeration, and backward masking with change detection

Factor	<i>df</i> 1	<i>df</i> 2	<i>F</i>	<i>p</i> <	$\eta_p^2$
Forward Masking–Enumeration					
SOA	3	45	94.912	.001	.864
Set Size	1.749	26.230	49.475	.001	.767
SOA × Set Size	5.576	83.643	1.710	.135	.102
Forward Masking–Change Detection					
SOA	3	45	9.222	.001	.381
Set Size	2	30	63.330	.001	.809
SOA × Set Size	6	90	2.392	.035	.138
Backward Masking–Enumeration					
SOA	1.529	22.935	56.031	.001	.789
Set Size	1.530	22.958	64.574	.001	.811
SOA × Set Size	3.801	57.020	12.490	.001	.454
Backward Masking–Change Detection					
SOA	3	45	34.722	.001	.698
Set Size	2	30	130.376	.001	.897
SOA × Set Size	6	90	2.784	.017	.157

For each of a condition's main and interaction effects, the degrees of freedom of the numerator, the degrees of freedom of the denominator, the *F* value, the significance level, and the goodness of fit of the general linear model are displayed.

Increasing the SOA between the forward mask and the to-be-enumerated items altered performance within the subitizing range (mean(1–4 items), SOA 200 vs. 24 ms:  $t(15) = 9.939$ ,  $p < .001$ ). For all set sizes, performance reached a plateau by around 200 ms (mean(1–4 items), SOA 494 vs. 200 ms:  $t(15) = 1.161$ , n.s.; Fig. 5A). For change detection, this amelioration of performance with increasing SOA was only observable with two-item displays, and it continued up to the 494-ms SOA [two items at SOA 494 vs. 24 ms:  $t(15) = 5.338$ ,  $p < .001$ ]. Visual working memory for higher set sizes did not benefit extraordinarily from increased SOA (mean(4,6 items), SOA 494 vs. 24 ms:  $t(15) = 2.207$ , n.s.; Fig. 5B). This pattern of results suggests that the forward-masking procedure successively affected the individuation of multiple items, eventually limiting the consolidation of information into visual working memory at a very early level of visual processing.

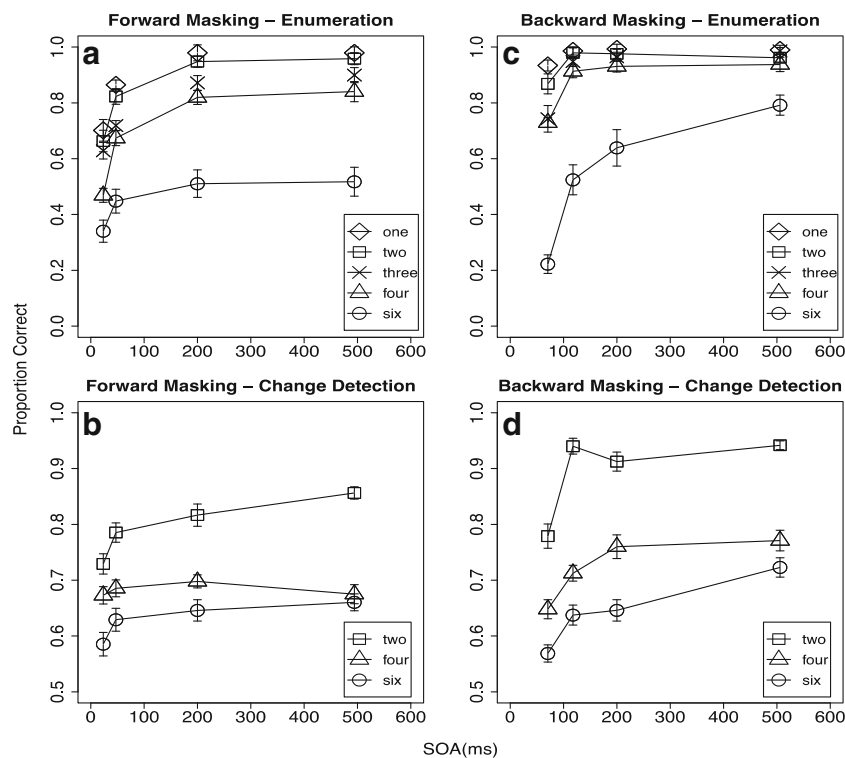
The results with forward and backward masking differed in two main ways. First, the forward-masking conditions had generally lower performance, perhaps due to the effect of the simultaneous mask. This simultaneous mask allowed us to study the time course of individuation by creating a limit on the degree to which features could be extracted for multiple objects simultaneously.

Second, the backward-masking effects were most noticeable with larger set sizes (six items for enumeration, or four items for change detection). Consistent with our hypothesis, this is particularly true within the time course of interruption masking (SOA > 100 ms). Within the subitizing range, increasing the SOA from 118 to 506 ms did not improve enumeration performance (mean(1–4 items), SOA 506 vs. 118 ms:  $t(15) = 1.464$ , n.s). However, for the larger set size (6 items), there was a significant improvement for longer SOAs [six items at SOA 506 vs. 118 ms:  $t(15) = 5.374$ ,  $p < .001$ ; Fig. 5C]. Similarly, in the case of change detection, we observed no benefit from larger SOAs in two-item displays [two items at SOA 506 vs. 118 ms:  $t(15) = 0.151$ , n.s.], while for four- and six-item displays, performance was better with the longest SOA (mean(4,6 items), SOA 506 vs. 118 ms:  $t(15) = 4.829$ ,  $p < .001$ ; Fig. 5D). Given the fact that backward masking had an effect at longer SOAs and larger set sizes, this is consistent with previous suggestions of a specific effect on the consolidation of object file content (Gegenfurtner & Sperling, 1993; Vogel et al., 2006).

For backward masking, only masks presented immediately after target offset (71-ms SOA), within the range of integration masking, influenced enumeration within the subitizing range (mean(1–4 items), SOA 118 vs. 71 ms:  $t(15) = 4.900$ ,  $p < .001$ ). For longer SOAs, however, enumeration performance was already at ceiling [see above for the nonsignificant effect of mean(1–4 items), SOA 506 vs. 118 ms; Fig. 5C). Together with the results of the forward masking, this pattern of results is consistent with the idea that subitizing is not instantaneous, but rather depends on the effective duration of the stimulus (Wutz et al., 2012). In a similar way, change detection performance for two-item displays was only altered by this very short SOA, and reached asymptote thereafter [two items at SOA 118 vs. 71 ms:  $t(15) = 6.203$ ,  $p < .001$ ; see above for the nonsignificant effect on two items at SOA 506 vs. 118 ms]. Thus, these results suggest that object identification can occur to a limited extent temporally in parallel with or very quickly after individuation. The typical four-item limit in visual short-term memory (Cowan, 2000; Luck & Vogel, 1994), however, is not reached within this very short period of time. Visual working memory measures for higher set sizes increased gradually with increasing backward-mask SOAs (see above for the significant effect at SOAs of 506 vs. 118 ms; Fig. 5D).

#### Visual masking and target detection

Neither the forward nor the backward mask showed the same dramatic reduction in performance for detection as had been found in the main experiment with individuation or identification. For both forward and backward masking, detection performance was above 90 % for almost all set



**Fig. 5** Results of the four crossed conditions in the main experiment: Forward masking with enumeration (panel A), forward masking with change detection (panel B), backward masking with enumeration (panel C), and backward masking with change detection (panel D). (A) Proportions of correct trials as a function of forward-mask SOA, for different item set sizes during enumeration. For reasons of better comparability to the change detection condition, only performance values for set sizes 2, 4, and 6 are connected with solid lines across the different SOAs. (B) Proportions of correct trials as a function of forward-mask SOA, for different item set sizes during change

detection. (C) Proportions of correct trials as a function of backward-mask SOA, for different item set sizes during enumeration. For reasons of better comparability to the change detection condition, only performance values for set sizes 2, 4, and 6 are connected with solid lines across different SOAs. (D) Proportions of correct trials as a function of backward-mask SOA, for different item set sizes during change detection. Error bars display one standard error of the mean for within-subjects designs. Individual performance values were centered on the mean performance of each subject before calculating the standard errors

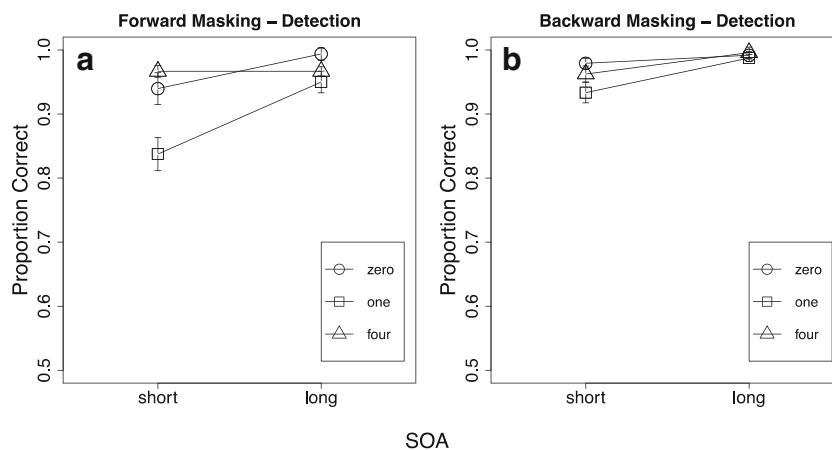
sizes and SOAs, in the cases of both correct rejections in target-absent trials (set size 0) and hits in target-present trials (Fig. 6). However, for both forward and backward masking, a significant effect of SOA was observable [forward masking,  $F(1, 9) = 17.778$ ,  $p < .002$ ,  $\eta_p^2 = .664$ ; backward masking,  $F(1, 9) = 10.494$ ,  $p < .01$ ,  $\eta_p^2 = .538$ ]. A major component of these effects was the worse performance for one-item displays at short SOAs [long vs. short SOA: forward masking and one item,  $t(9) = 3.017$ ,  $p < .03$ ; backward masking and one item,  $t(9) = 3.074$ ,  $p < .026$ ; see Fig. 6]. This pattern of results resembles that for enumeration performance under the influence of masking (Fig. 5A). For one-item displays, detection conceivably is also the main component of enumeration; therefore, it is reasonable that these two conceptually very similar conditions would yield comparable results under forward and backward masking. In other words, detection is a limiting factor in the enumeration of one-item displays.

In general, however, the average  $d'$  were high under all conditions (forward masking and short SOA,  $M = 3.588$ ,

$SD = 1.266$ ; forward masking and long SOA,  $M = 4.930$ ,  $SD = 0.924$ ; backward masking and short SOA,  $M = 4.347$ ,  $SD = 1.037$ ; backward masking and long SOA,  $M = 5.487$ ,  $SD = 1.013$ ). It is important to note that both forms of visual masking—forward and backward—yielded similar results: Target detection was not greatly affected by these masking techniques. This strikingly good detection performance contrasts with the significant masking effects on both enumeration and change detection, even though the same temporal parameters, in terms of SOA and visual stimuli, were used. These results are consistent with the control experiment reported in our recent study of rapid individuation, which also showed that forward and simultaneous masking did not simply reduce target visibility indiscriminately (Wutz et al., 2012).

A second critical difference between the results of the control study and those of the main experiment is that the worst performance was found with one-target displays under forward masks with short SOAs, as compared to performance with four items. This is the opposite trend from the





**Fig. 6** Results of the control conditions: forward masking with detection (panel A) and backward masking with detection (panel B). (A) Proportions of correct trials as a function of forward-mask SOA for different item set sizes during detection. (B) Proportions of correct trials as a function of backward-mask SOA for different item set sizes

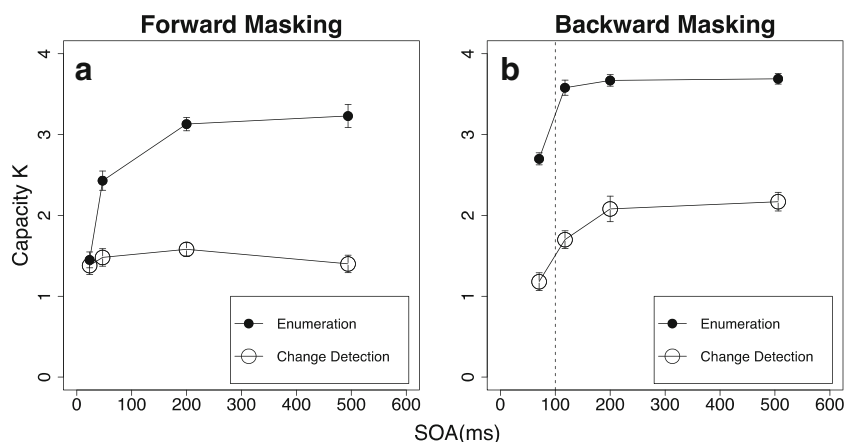
during detection. Error bars display one standard error of the mean for within-subjects designs. Individual performance values were centered on the mean performance of each subject before calculating the standard errors

enumeration conditions in the main experiment, in which performance was better for one item than for four. Displays with one item were harder to detect than those with higher set sizes [for the 24-ms SOA with one vs. four items:  $t(9) = -5.127, p < .001$ ], whereas in the main experiment, smaller set sizes were easier to enumerate than were higher numerosities [for the 24-ms SOA with one vs. four items:  $t(15) = 4.751, p < .001$ ]. As target detection either was not affected at all by masking or showed the reverse pattern of results, as compared to enumeration, the powerful effects of visual masking on the formation of object files reported above cannot be explained by a failure to register the presence of a target display. Instead, the reported results reveal distinct

effects of integration and interruption masking on the extraction of object-like representations from the sensory signal after it has already been registered by the observer as new input, reflecting temporal limits on the perceptual computations *within* the sensory image for the time of its persistence.

Visual masking and object capacity

In order to better understand the accumulation of object information over time, within and beyond the period of visual persistence, we compared object capacity estimates (see the *Method* section) across the four conditions (Figs. 3 and 7). Consistent with a recent study, capacity limits were



**Fig. 7** Results of the four crossed conditions in the main experiment, translated into the capacities *K* for forward masking (panel A) and backward masking (panel B). (A) Capacity *K* as a function of forward-mask SOA for enumeration (filled circles) and change detection (open circles). (B) Capacity *K* as a function of backward-mask SOA for enumeration (filled circles) and change detection (open circles). The vertical dashed line indicates the temporal window of visible

persistence (Di Lollo, 1980), where the influence of backward masking switches from integration to interruption masking (Scheerer, 1973; Scheerer & Bongartz, 1973). Error bars display one standard error of the mean for within-subjects designs. Individual performance values were centered on the mean performance of each subject before calculating the standard errors

higher for the enumeration task than for the visual working memory task (Piazza et al., 2011). Of particular interest, however, are the temporal dynamics of these capacity differences, showing a clear dissociation between forward/integration and backward/interruption masking in the two tasks. Whereas enumeration capacity increased throughout the whole time course of the forward-masking procedure, backward masking influenced enumeration only at the very short SOA immediately after target offset (in the time period of integration masking). Visual working memory (i.e., change detection) capacity, however, did not increase as a function of forward-mask SOA (staying flat at around 1.5 items), but rose gradually with longer SOAs to the backward mask, up to more than two items (Fig. 7).

This reasoning is confirmed by a within-subjects ANOVA on the capacity estimates for the two tasks within the respective time courses of integration and interruption masking. The applied forward-masking technique was specifically designed to vary integration masking. For backward masking, however, a distinction between short (below 100 ms) and long SOAs has to be made (Scheerer, 1973; Scheerer & Bongartz, 1973). While integration masking is likely to occur for short SOAs, masks with a longer SOA to the target display have an interrupting influence on visual processing. A trend test on linearity for the capacity estimates for enumeration throughout the forward-masking SOAs revealed a significant effect [ $F(1, 15) = 82.989$ ,  $p < .001$ ,  $\eta_p^2 = .847$ ], whereas no such linear trend was observable for the memory task within the same temporal range [ $F(1, 15) = 0.088$ , n.s.,  $\eta_p^2 = .006$ ]. In contrast, after the time period at which backward masking has an interrupting influence on the perceptual process (around 100 ms; Scheerer, 1973; Scheerer & Bongartz, 1973), only visual working memory capacity increased linearly with longer SOAs [ $F(1, 15) = 8.222$ ,  $p < .015$ ,  $\eta_p^2 = .354$ ]. Enumeration capacity, however, had already reached asymptotic values by 100 ms and showed no further linear effects [ $F(1, 15) = 0.877$ , n.s.,  $\eta_p^2 = .055$ ].

In order to pin down this interaction between task and masking type statistically, we calculated the average performance increase in terms of capacity from the shortest to the longest SOA within the respective time courses of integration (forward) masking (24- to 494-ms SOAs) and interruption (backward) masking (118- to 506-ms SOAs):

$$\Delta K_{\text{for}} = K_{\text{SOA}494\text{ms}} - K_{\text{SOA}24\text{ms}}, \quad (1)$$

$$\Delta K_{\text{back}} = K_{\text{SOA}506\text{ms}} - K_{\text{SOA}118\text{ms}}. \quad (2)$$

These capacity differences were subject to a within-subjects ANOVA with the factors Task (enumeration, change detection) and Masking Type (forward, backward). Both main effects were significant [task,  $F(1, 15) = 17.207$ ,

$p < .001$ ,  $\eta_p^2 = .534$ ; masking type,  $F(1, 15) = 11.888$ ,  $p < .004$ ,  $\eta_p^2 = .442$ ]. More importantly, both factors interacted [Task  $\times$  Masking Type,  $F(1, 15) = 66.861$ ,  $p < .001$ ,  $\eta_p^2 = .817$ ]: Whereas enumeration capacity increased over time under forward/integration masking [ $M(\Delta K_{\text{for}}) = 1.78$  items,  $SD = 0.78$  items;  $M(\Delta K_{\text{back}}) = 0.12$  items,  $SD = 0.50$  items], vSTM capacity increased over time with backward/interruption masking [ $M(\Delta K_{\text{for}}) = 0.02$  items,  $SD = 0.71$  items;  $M(\Delta K_{\text{back}}) = 0.47$  items,  $SD = 0.65$  items].

Summing up, the type of visual masking interacted with the task performed. Interruption masking appeared to exclusively influence the consolidation of information in visual working memory, with little effect on enumeration. Conversely, increasing the forward-mask SOA yielded gradually increasing capacity in the enumeration task, while change detection capacity remained stably poor throughout the whole range of SOAs.

## Discussion

Overall, the findings are consistent with the hypothesized effect of masking on different stages of object processing (Fig. 1). These results suggest a close link between capacity limits (in both subitizing and visual working memory) and temporal constraints on object individuation and identification. It adds to extensive empirical and theoretical work that has indicated that object file formation involves a temporal succession of processing steps: Target detection is faster than target identification in visual search (Sagi & Julesz, 1985); postoffset location information is processed sooner than identity information (Finkel & Smythe, 1973; Schiller, 1965); spatiotemporal information allows an “object file” to be created, before it is filled in with object features (Kahne-man et al., 1992); and spatial locations are preattentively indexed first, followed by featural information only becoming available later to attention-dependent mechanisms (Pylyshyn, 1994).

This raises the question of why different spatiotemporal windows are involved in object perception, one reflecting individuation (visual persistence) and one limiting identification (consolidation into vSTM). One possible explanation is that this situation reflects the brain’s strategy to deal with the need to spatially and temporally integrate information coming from a continuous flow of sensory information.

As is known from mathematical and engineering sciences, nonlinear positive and/or delayed feedback systems that are engaged in real-time processing exhibit asymptotic unstable behavior when confronted with signals with different latencies that have to be combined (Sandberg, 1963). In such a system, there is a disequilibrium between the need for dynamic and flexible representations (emphasizing new

information) and the need for stable and reliable visual representations (maintaining the current state). This trade-off between stimulus read-out and perceptual synthesis can be achieved by temporal multiplexing of feedforward and feedback signals (Öğmen, 1993).

We suggest that this need to balance feedforward and feedback processes must inherently limit capacity for rapid object individuation. According to this model, the real-time dynamics of visual processes unfold in three phases: (1) Afferent feedforward signals allow read-out of the sensory information; (2) during the decay of the feedforward signal, a feedback or reentrant dominant phase establishes perceptual synthesis; (3) a reset phase is initiated, resulting in an inhibition of the feedback signals and a reestablishment of the feedforward-dominant mode that delivers the new signal. This succession of transient epochs implements a degree of inertia in the system's response to changes in input, and thus limits its real-time dynamics in order to guarantee an equilibrium between the flexibility and stability of the visual representation (Enns & Di Lollo, 2000; Öğmen, 1993). Of course, the solution of multiplexing creates temporal windows of visual persistence, during which only a limited number of objects can be processed.

In accordance with this idea, we have reported evidence that capacity limits in enumeration depend, at least in part, on a “magic window” of sensory persistence (see also Wutz et al., 2012) that determines the “magic number” of around 4. Using integration masking, the effective persistence of the target display can be fractionated (Di Lollo, 1980; Wutz et al., 2012), thus reducing the effective lifespan of the feedforward dominant phase, and thus limiting the time to read-out spatiotemporal object information and create “object files.” This forward-masking technique appears to act early in the individuation stage in which targets are segmented and spatiotemporally segregated from the background.

As we described above, in a second phase the effective signal strength decays, and the system enters the reentrant phase of processing, during which object identification mechanisms fill in the feedforward-established “object files” with featural content. Thus, in addition to the first capacity limit resulting from the effective persistence of the stimulus, a secondary limit comes from the consolidation of information into visual working memory. In particular, this consolidation process can be interrupted if new visual input arrives during the phase of feedback identification processing, since this new stimulus initiates a new feedforward process during this crucial phase of inertia, leading to “interruption masking” (Breitmeyer & Öğmen, 2006; Enns & Di Lollo, 2000; Scheerer, 1973; Scheerer & Bongartz, 1973). The speed of object identification, and thus the formation of high-resolution object files, is influenced by processing demands and encoding complexity (Alvarez & Cavanagh,

2004). As demonstrated previously, visual working memory performance rises gradually to asymptote under the influence of long backward masks (Gegenfurtner & Sperling, 1993; Vogel et al., 2006). Consequently, the overall capacity of vSTM is limited by temporal buffering, both at the feedforward individuation and the reentrant identification stages of object processing.

The dependence of object individuation capacity on the time window of temporal integration and visual persistence further fosters the central role that individuation can play in mediating between the two opposing needs of the visual system in real-time processing: flexibility and stability. The fixed number of newly established object files is a direct consequence of the time period of initial feedforward processing, which is a fundamental and computationally inherent characteristic of the temporal dynamics of visual processing. The information gathered during this constant temporal window enables the organism to preserve basic behavioral potentials, such as reacting to spatiotemporal changes in the environment by body or eye movements. In order to achieve more sophisticated interaction with the environment (like identification or memory), and therefore stability on a higher representational resolution, additional processing is necessary, at the cost of flexibility to new input.

Although human cognition is remarkably powerful, its online workspace, working memory, appears to be highly limited in the number of informational units it can process (Cowan, 2000; Luck & Vogel, 1997; Sperling, 1960). Here we provide a specific and experimentally testable hypothesis for the origin of cognitive capacity limitations: processing time. Previous proposals about the root of capacity limitations in vision have introduced relatively abstract concepts like “slots” (Fukuda, Awh, & Vogel, 2010; Luck & Vogel, 1997) or “resources” (Alvarez & Cavanagh, 2004; Bays & Husain, 2008). While these theories clearly have augmented our understanding of visual object capacity on a descriptive level, the present explanation for capacity limits accounts for them in terms of known mechanisms and embeds the ongoing debate about processing limits in the already well-established body of work about the temporal dynamics of the visual system (Busch, Dubois, & VanRullen, 2009; Enns & Di Lollo, 2000; Gegenfurtner & Sperling, 1993; Loftus et al., 1992; Roelfsema, Lamme, & Spekreijse, 2000; Shallice, 1964; Singer, 1999; Sperling, 1960, 1963; Ullman, 1984; VanRullen & Koch, 2003; Wundt, 1899, 1900). As was also stated above, our explanation is fully compatible with resource- or slot-based approaches, but it emphasizes a different perspective on the formation of object representations that can be empirically investigated and directly observed in the laboratory.

In practical terms, such an approach would allow for normal or clinically relevant variability in processing capacity to be broken down into concrete factors, such as

variations in temporal integration periods, speeded-up or slowed-down employment of selective attention, or altered read-out slopes of individuation mechanisms. On a theoretical level, we argue that formal descriptions of selective attention and object file formation (Blaser, Sperling, & Lu, 1999; Itti, Koch, & Niebur, 1998; Koch & Ullman, 1985) should be augmented by a temporal dimension and not solely focus on spatial characteristics of the visual display (Burr, 1984; Burr, Ross, & Morrone, 1986; Dempere-Marco et al., 2012; Lisman & Idiart, 1995). The explanation of object capacity in terms of temporal constraints on the underlying mechanisms fosters the link between space and time, as well as the role of both of these a priori concepts in sensation (Kant, 1899). These two aspects are both fundamental to human cognition, since “space and time are the pure forms of . . . [sensation]” (Kant, 1899, p. 164; a change to the original by A.W. is indicated by the square brackets).

**Author note** The research was supported in part by European Research Council Grant (Grant Agreement n. 313658); by the Italian Ministry of Universities and Research (MIUR), Project No. PRIN 2009; and by the Fondazione Cassa di Risparmio di Trento e Rovereto. We thank Jennifer Corbett and Alfonso Caramazza for helpful comments on an earlier version of this article.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

## 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. doi:10.1111/j.0963-7214.2004.01502006.x
- Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science*, *321*, 851–854. doi:10.1126/science.1158023
- Blaser, E., Sperling, G., & Lu, Z.-L. (1999). Measuring the amplification of attention. *Proceedings of the National Academy of Sciences*, *96*, 11681–11686.
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436. doi:10.1163/156856897X00357
- Breitmeyer, B. G., & Ögmen, H. (2006). *Visual masking: Time slices through conscious and unconscious vision* (2nd ed.). Oxford, UK: Oxford University Press.
- Burr, D. C. (1984). Summation of target and mask metacontrast stimuli. *Perception*, *13*, 183–192.
- Burr, D. C., Ross, J., & Morrone, M. C. (1986). Seeing objects in motion. *Proceedings of the Royal Society B*, *227*, 249–265. doi:10.1098/rspb.1986.0022
- Busch, N. A., Dubois, J., & VanRullen, R. (2009). The phase of ongoing EEG oscillations predicts visual perception. *Journal of Neuroscience*, *29*, 7869–7876.
- Cowan, N. (2000). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*, 87–114. doi:10.1017/S0140525X01003922. disc. 115–185.
- Dempere-Marco, L., Melcher, D. P., & Deco, G. (2012). Effective visual working memory capacity: An emergent effect from the neural dynamics in an attractor network. *PLoS ONE*, *7*, e42719. doi:10.1371/journal.pone.0042719
- Di Lollo, V. (1980). Temporal integration in visual memory. *Journal of Experimental Psychology: General*, *109*, 75–97. doi:10.1037/0096-3445.109.1.75
- Enns, J. T., & Di Lollo, V. (2000). What’s new in visual masking? *Trends in Cognitive Sciences*, *4*, 345–352. doi:10.1016/S1364-6613(00)01520-5
- Finkel, D. L., & Smythe, L. (1973). Short-term storage of spatial information. *Developmental Psychology*, *9*, 424–428.
- Fukuda, K., Awh, E., & Vogel, E. K. (2010). Discrete capacity limits in visual working memory. *Current Opinion in Neurobiology*, *20*, 177–182.
- Gegenfurtner, K. R., & Sperling, G. (1993). Information transfer in iconic memory experiments. *Journal of Experimental Psychology: Human Perception and Performance*, *19*, 845–866. doi:10.1037/0096-1523.19.4.845
- Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*, 1254–1259.
- Jevons, W. S. (1871). The power of numerical discrimination. *Nature*, *3*, 363–372.
- Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, *24*, 175–219. doi:10.1016/0010-0285(92)90007-O
- Kant, I. (1899). Kritik der reinen Vernunft. In R. Schmidt (Ed.), *Kritik der reinen Vernunft*. Hamburg, Germany: Meiner.
- Kaufman, E. L., Lord, M. W., Reese, T., & Volkman, J. (1949). The discrimination of visual number. *The American Journal of Psychology*, *62*, 496–525.
- Klein, S. E. (2001). Measuring, estimating, and understanding the psychometric function: A commentary. *Perception & Psychophysics*, *63*, 1421–1455.
- Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology*, *4*, 219–227.
- Lisman, J. E., & Idiart, M. A. (1995). Storage of  $7 \pm 2$  short-term memories in oscillatory subcycles. *Science*, *267*, 1512–1515.
- Loftus, G. R., Duncan, J., & Gehrig, P. (1992). On the time course of perceptual information that results from a brief visual presentation. *Journal of Experimental Psychology: Human Performance and Perception*, *18*, 530–549. doi:10.1037/0096-1523.18.2.530
- Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279–281. doi:10.1038/36846
- Melcher, D., & Piazza, M. (2011). The role of attentional priority and saliency in determining capacity limits in enumeration and visual working memory. *PLoS ONE*, *6*, e29296. doi:10.1371/journal.pone.0029296
- Neisser, U. (1967). *Cognitive psychology*. New York, NY: Appleton-Century-Crofts.
- Ögmen, H. (1993). A neural theory of retino-cortical dynamics. *Neural Networks*, *6*, 245–273.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442. doi:10.1163/156856897X00366
- Phillips, W. A., & Baddeley, A. D. (1971). Reaction time and short-term visual memory. *Psychonomic Science*, *22*, 73–74.
- Piazza, M., Fumarola, A., Chinello, A., & Melcher, D. (2011). Subitizing reflects visuo-spatial object individuation capacity. *Cognition*, *121*, 147–153.
- Pylyshyn, Z. (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. *Cognition*, *32*, 65–97. doi:10.1016/0010-0277(89)90014-0

- Pylyshyn, Z. (1994). Some primitive mechanisms of spatial attention. *Cognition*, *50*, 363–384.
- Roelfsema, P. R., Lamme, V. A. F., & Spekreijse, H. (2000). The implementation of visual routines. *Vision Research*, *40*, 1385–1411.
- Sagi, D., & Julesz, B. (1985). Enhanced detection in the aperture of focal attention during simple discrimination tasks. *Nature*, *321*, 693–695.
- Sandberg, L. W. (1963). Signal distortion in nonlinear feedback systems. *Bell System Technical Journal*, *42*, 2533–2550.
- Scheerer, E. (1973). Integration, interruption and processing rate in visual backward masking. *Psychologische Forschung*, *36*, 71–93. doi:10.1007/BF00424655
- Scheerer, E., & Bongartz, W. (1973). Integration, interruption and processing rate in visual backward masking. *Psychologische Forschung*, *36*, 95–115. doi:10.1007/BF00424965
- Schiller, P. H. (1965). Backward masking for letters. *Perceptual and Motor Skills*, *20*, 47–50.
- Shallice, T. (1964). The detection of change and the perceptual moment hypothesis. *British Journal of Statistical Psychology*, *17*, 113–135.
- Singer, W. (1999). Time as coding space? *Current Opinion in Neurobiology*, *9*, 189–194.
- Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied*, *74*(11, Whole No. 498), 1–29.
- Sperling, G. (1963). A model for visual memory tasks. *Human Factors*, *5*, 19–31.
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*, 97–136. doi:10.1016/0010-0285(80)90005-5
- Ullman, S. (1984). Visual routines. *Cognition*, *18*, 97–159.
- VanRullen, R., & Koch, C. (2003). Is perception discrete or continuous? *Trends in Cognitive Sciences*, *7*, 207–213.
- Vogel, E. K., Woodman, G. F., & Luck, S. J. (2006). The time course of consolidation in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 1436–1451. doi:10.1037/0096-1523.32.6.1436
- Wundt, W. (1899). Zur Kritik tachistoskopischer: Versuche I. *Philosophische Studien*, *15*, 287–317.
- Wundt, W. (1900). Zur Kritik tachistoskopischer: Versuche II. *Philosophische Studien*, *16*, 61–81.
- Wutz, A., Caramazza, A., & Melcher, D. (2012). Rapid enumeration within a fraction of a single glance: The role of visible persistence in object individuation capacity. *Visual Cognition*, *20*, 717–732.
- Xu, Y., & Chun, M. M. (2009). Selecting and perceiving multiple visual objects. *Trends in Cognitive Sciences*, *13*, 167–174. doi:10.1016/j.tics.2009.01.008