Distinct neural mechanisms meet challenges in dynamic visual attention due to either load or spatial interference

Mäki-Marttunen V.[1,], Hagen T.[1,*], Laeng, B.[1,2], Espeseth, T.[1,3]*

[1] Department of Psychology, University of Oslo, Oslo, Norway.
* corresponding authors: makimarttunen.veronica@gmail.com; tmshag1@gmail.com. Address: Postboks 1094 Blindern 0317 Oslo, Norway. T.E. +47-93653834
[2] *RITMO Centre for Interdisciplinary Studies in Rhythm, Time and Motion,* University of Oslo, Oslo, Norway
[3]Norwegian Centre for Mental Disorders Research (NORMENT), KG Jebsen Centre for Psychosis Research, Division of Mental Health and Addiction, Oslo University Hospital, Norway

Abstract

When solving dynamic visuo-spatial tasks, like multiple-object tracking (MOT), the brain copes with challenges imposed by the number of objects (i.e. load) that has to be tracked, but also by spatial interference from irrelevant objects (i.e. crowding). At present, it is not clear whether load and crowding activate separate cognitive and physiological mechanisms. In particular, cognitive workload recruits cortical systems, like the dorsal and ventral attention networks, and brainstem neuromodulatory systems, such as the locus coeruleus - norepinephrine (LC-NE) system and the ventral tegmental area/substantia nigra dopaminergic (VTA/SN - DA) system. However, it remains unknown what the specific roles of each of these brain systems are in dealing with load versus crowding. We employed a MOT task, where participants had to track simultaneously multiple randomly moving targets. We varied the degree of processing challenge in two ways: first, how many dynamic targets participants had to track ("load", known at the beginning of every trial); second, the proximity between targets and distractors (i.e., the number of close target-distractor encounters, causing moment-to-moment confusion within the trial). To evaluate the extent to which load and crowding elicits distinct physiological responses, we ran two MOT experiments in separate cohorts. In a first experiment, we measured task-induced saccades and pupil dilations. In a separate fMRI experiment, we measured brain activity elicited by MOT. The behavioral results from both experiments revealed that increased load and crowding led to reduced accuracy in an additive manner, suggesting that they may rely on different mechanisms. Consistent with this, analysis of the ocular data showed that load was associated with pupil dilations, whereas crowding was not. Conversely, both load and crowding were associated with increased frequency of saccades during tracking. The fMRI results showed that dorsal attentional areas were proportionally active with both higher levels of load and crowding. Higher crowding recruited additionally ventral attentional areas that may reflect orienting mechanisms. The activity in the brainstem nuclei VTA/SN and LC showed strongly dissociated patterns. While activations in VTA/SN mainly reflected variations in crowding, activity in LC mainly reflected variations in load. An exploratory individual differences' approach showed that high-performers were able to recruit dynamically the LC with high difficulty. Thus, distinct brain and brainstem systems deal with spatially versus non-spatially-related effort.

Key words: visual attention, multiple-object tracking, cognitive load, crowding, fMRI, pupillometry

1. Introduction

When faced with a visuo-spatial task that requires top-down control -that is, a task that cannot be solved solely automatically-, the logistics of attention become engaged. The brain continuously assesses the appropriate level of processing resources and recruit optimally the mechanisms that deal with current demands. In the case of multiple-object tracking (MOT, (1), Figure 1.A), the brain flexibly prepares for the attentional challenge posed by the number of objects to be tracked simultaneously ("load"; (2, 3)). Subsequently, the brain copes with emerging, unpredicted situations of spatial interference during the tracking period. Specifically, factors that affect accuracy in MOT are the speed of the objects (4, 5), the duration of tracking (6), and the frequency of close encounters between targets and distractors (a concept known as "crowding", (7-12)).

All these factors seem to be associated with the degree of proximity between targets and distractors over a certain tracking period. That is, both increased speed of movement and increased load would on average lead to a higher frequency of close encounters within one tracking period, suggesting that crowding could be the factor that explains all the challenges mentioned above (13). Hence, a still open question regarding the MOT paradigm was whether the deterioration of performance could be simply reduced to spatial interference, that is, to a limitation in attentional resolution (14). Two approaches would be required to address this question directly. First, a task where both load and crowding are manipulated so that any unique or interactive effects may be scrutinized. Second, measures ocular and brain activity signatures that reveal whether specific physiological responses are associated with each type of challenge.

Recent neuroimaging research has revealed that a fronto-parietal brain network is recruited during the tracking of multiple objects, showing increased activity with increasing load (15-21). This is consistent with findings that implicate this network in dealing with attentional demands across tasks and modalities (22-24). In addition, neuromodulatory systems, such as the locus coeruleus-norepinephrine (LC-NE) and ventral tegmental area/substantia nigra (VTA/SN) systems, enable coping with visual-attentional and cognitive challenges (25). In MOT, the LC-NE system is recruited with higher load (17). However, a still open question is how these systems participate in resolving the spatial interference that may arise during tracking. In particular, it would be expected that both the number of objects to track and the events of crowding should engage an effort-related system, but it remains unclear what is the specificity of this system. One possibility is that load and crowding engage the same cognitive mechanisms; for example, that they both relate to the ability of keeping differentiated mental representations of the objects (2, 13). In this case, manipulating either load or crowding would recruit the same cortical and subcortical regions. Another possibility is that the nature of the limitation that load and crowding pose is different, and then we would expect a distinction at the neural level. For example, while a general demand network would be engaged parametrically with load to sustain and update the representation of the objects, additional resources may help dealing with unexpected events of spatial interference. This would imply a more fine-regulated processing that provides more specificity and efficacy. The cortical and neuromodulatory activity associated to simultaneous spatial and non-spatial challenges have never been tested directly, to the best of our knowledge. Since load and crowding can be manipulated independently in MOT, we can elucidate between these possibilities.

We adapted the MOT task by manipulating load and spatial interference (amount of crowding situations that occurred along the tracking period). The load was known to the participants at the assignment of the objects to track. In addition, we varied the amount of crowding during the trials, with trials of low or high crowding. Importantly, in the high crowding condition the number of close encounters between targets and distractors was equalized across load levels; that is, there

was the same number of close encounters in the 2, 3 and 4 targets load levels. We hypothesized that, if load and spatial interference due to crowding are actually the same kind of spatial limitation, keeping constant the amount of spatial interference at the different load levels would warrant equal accuracy (Figure 1.f). If load and crowding rely on partly different processes, then increasing the level of load should lead to decreases in accuracy at both crowding levels (Figure 1.g). On the other hand, if load and crowding share mechanistic factors, we would expect that at equal levels of one factor, there would be different patterns of accuracy for the other (Figure 1.h, (11)). This would be expected because the two loading factors would act as "dual-tasks" (26, 27). A within-subject design would help disentangle these possibilities.

We measured physiological signatures of the concurrent manipulations. In two separate studies with different cohorts, ocular responses (saccades and pupil dilations) and brain activity (cortical and brainstem) were acquired. Pupil dilations have classically been thought to provide a reliable signature of mental effort (28-31), either as direct engagement or decision to engage ((30, 32)), and have been strongly tied to the LC-NE system. Saccades have been shown to be related to the VTA/SN-DA system. In the second study, brain imaging allowed us to study whether the same or different brain circuits are involved at varying levels of load and crowding. We studied both cortical and brainstem activity, focusing on two brainstem nuclei: locus coeruleus and VTA/SN. The LC projects to the whole cortex and facilitates visuo-spatial attention (25, 33-35). Neuroimaging studies in humans show that LC activity is related to pupil size (36) and that both LC activity and pupil changes reflect the increasing effort posed by load in MOT (17). VTA/SN projects to frontal and subcortical areas and participates in motivated and effortful behavior (37). Importantly, the flexible recruitment of these systems may at least partially underlie individual differences in performance (38, 39). We pursued the following questions: first, how is accuracy influenced by the interaction of load and crowding (Figure 1)? Second, given that the pupil is sensitive to increasing demands due to load (17, 40), is it sensitive to increases of demands due to higher spatial interference as well? Is that the case with saccades? Third, are the same brain circuits involved under higher levels of load and spatial interference? Fourth, do the brainstem nuclei contribute differently to the tracking of multiple objects under the different conditions? Finally, we tested whether individual differences in performance can be associated to distinct patterns of recruitment of the neuromodulatory systems.

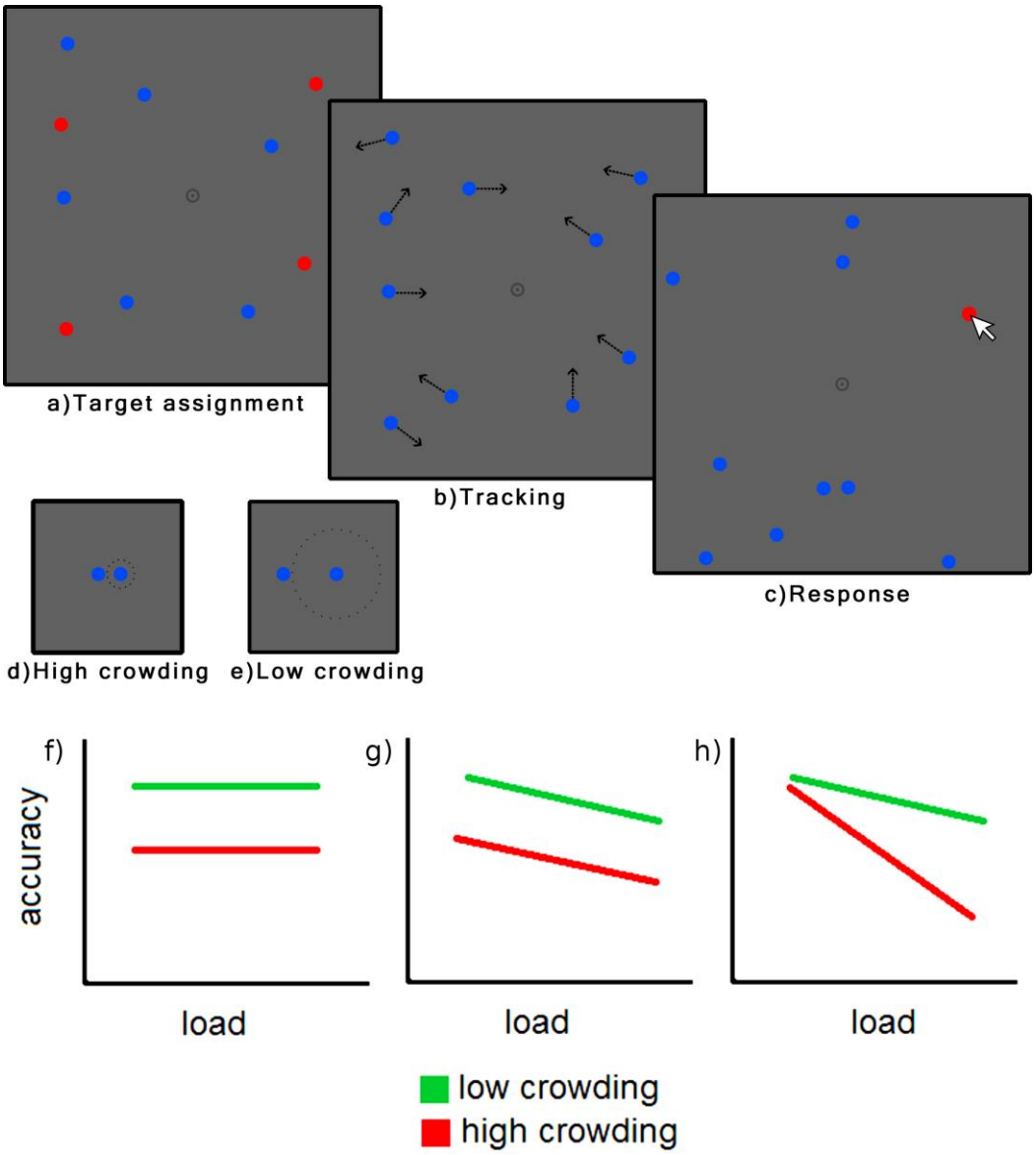

Figure 1. MOT task and hypothesized effects of load and crowding on accuracy. Top panel: MOT task. In the assignment period of a trial (a), ten black disks are presented. Then, a number of them (the targets) is highlighted in red, so as to obtain different load levels (i.e., 2, 3 or 4). Then, all objects turn black and start moving (b). Participants are instructed to track the target objects while they move. At the end of the tracking period (10 seconds), participants indicate which are the targets by clicking on them with the mouse (c). Crowding (i.e. spatial interference) was varied as the minimum distance at which distractors approached targets: >80 pixels in low crowding (e) and >15 pixels in high crowding (d). Bottom panel: Hypothesis. If in each crowding level (low/high) the amount of crowding (i.e. close encounters) is held constant across load levels, the hypothesized patterns differ whether load and crowding rely on the same or different mechanisms: f) expected pattern for the load effect being being given by the amount of crowding; g) expected pattern for a load-crowding model with independent mechanisms; h) expected pattern for a load-crowding model with shared mechanisms.

2. Materials and Methods

2.1 *MOT task*

In the MOT task, participants are presented with 10 objects (disks). Some of them are cued as targets by changing their color for a short period. Afterwards, these targets acquire the same appearance as the other objects (or distractors) and then all the disks start moving in random directions, changing paths only when bouncing with each other and with the borders of the display area. The participant's task is to track the targets during the tracking interval and, at the end of the trial, indicate which objects in the display belonged to the target set. We parametrically varied the number of target objects, or cognitive load, so that the participants should correspondingly adjust

the required degree of attentional effort. In the present experiments, we used three levels of load, with trials of two, three or four objects to be tracked. In addition, we included two levels of crowding for each of the load levels: low and high crowding. Crowding is defined as the number of instances where one distractor approaches a target within a certain close-range distance (see description below). This proximity increases confusion about the target, therefore increasing the risk of losing the target or swapping the target with a distractor.

The size of the display was 800x800 pixels (20.9 x 20.9 degrees of visual angle). Objects had a radius of 15 pixels (0.4 degrees). Trial duration was 10 seconds. The displays were redrawn at a rate of 30 frames per second and objects moved with a speed of 10 pixels per frame (0.27 degrees). We generated two sets of stimuli. In the "low crowding" trials the distance between targets and distractors that was always larger than 80 pixels (2.1 degrees; Figure 1.A). In the "high crowding" stimulus set, we allowed the distance between targets and distractors to be as short as 15 pixels (0.4 degrees). Close encounters in the high crowding trials were thus defined as situations where a distractor traveled within 15 and 60 pixels of a target (0.4 - 1.6 degrees) and kept a distance of less than 80 pixels for no more than 10 frames (333 ms). Thus, close encounters were brief, in order to ensure that some close encounters would not last over several seconds. Each load level in the "high crowding" set had exactly 13 close encounters, irrespective of load. In addition, it was ensured that each target had at least one close encounter with a distractor, but not more frequently than the 75% percentile of close encounter distributions in the generated trials. As close encounters could arguably still occur in the low crowding condition, we attempted to assert some control on the number of "long-range" close encounters. That is, we defined long-range close encounters as starting when a distractor approached a target closer than 100 pixels and ended when getting further away than 120 pixels. In selecting trials we ensured that the total number of long-range close encounters per trial would be less than the 75% percentile of such long-range close encounters in the generated set of trials prior to selection. Equalizing such long-range close encounters was not possible across load levels.

In summary, we had a mixed design with factors load (3 levels) and crowding (2 levels), making in total, six different conditions.

2.2 *Eye-tracking study*

2.2.1 *Participants*

Forty-one adults (28 females) were recruited through social media to participate in the study (mean age: 24.2; range: 19-40). They were compensated with a gift card equivalent to 100 Norwegian Kroner (NOK) per hour and each session took approximately 1 hour. The study was conducted according to institutional guidelines, and was approved by the local ethics committee (IRB).

2.2.2 *Experimental procedure*

The MOT task was written in code and delivered using custom software written in JavaScript. The participants read the instructions and performed a training session on a computer (10 trials). Afterwards, the full experiment was run in the lab. The total number of objects in the display was kept constant at 10. Each trial consisted of the presentation of the 10 objects in blue for 2 s. Then some of the disks -the targets- turned red for 2.5 s and then turned back to blue. In the passive viewing condition, no ball turned red. The disks started to move and the tracking period lasted 10 s (tracking trials) or 7 s (passive viewing trials). At the end of the tracking period all objects' movements stopped. Participants had to click on all the disks that were originally selected to be tracked. Participants clicked with the right button if they had high confidence on their response

and with the left button if they had low confidence. Response accuracy and time to start responding were recorded for each trial.

2.2.3 *Pupillometry analysis*

Blinks were removed and interpolated before hampel and lowess filtering were applied. For each blink detected by the SMI software, we linearly interpolated from 5 samples before onset until 5 samples after offset. Trials were baseline corrected using a baseline of 500 msecs before the beginning of the tracking period. The pupillometry data was analyzed in the time region 2633ms to 12000ms (determined by Monte Carlo analysis on load levels). To further explore whether the effect of crowding would become evident by the end of the trial, we analyzed the pupil change in two intervals, one early (2633 – 4633 msecs) and one late (10000 – 12000 msecs) within the trial. Saccades were detected with the SMI software and extracted during the tracking interval (2000-12000 ms).

2.3 *fMRI study*

2.3.1 *Participants*

Forty-one adults (25 females) were recruited through social media to participate in the study (mean age: 25; range: 19-37). They were compensated with a gift card equivalent to 100 Norwegian Kroner (NOK) per hour and each session took approximately 2 hours.

Before starting testing, participants were asked to answer a brief questionnaire to assess whether they passed the inclusion criteria (no serious neurological or psychiatric illness, vision or language problems, etc.). Participants with an average proportion of accuracy of < 0.6 were removed since this level of performance approaches chance level. The study was conducted according to institutional guidelines and approved by the local ethics committee.

2.3.2 *Experimental procedure*

The MOT task was included the same stimulus set as the eye-tracking study and was delivered using E-prime 2. The participants read the instructions and performed a training session before the fMRI session (10 trials). Different from the eye-tracking study, in the fMRI study we used partial report. At the end of the tracking period all objects' movements stopped and one of the objects was highlighted with a yellow circle. The participant responded whether the probed object was one of the targets that was originally selected to be tracked. The response window was 2 s, after which the trial ended and was followed by a 4-s inter-trial interval (ITI) that consisted of a fixation cross. Response accuracy and reaction time were recorded for each trial. The probability that the probed object was one of the targets was 50% for all tracking conditions.

2.3.3 *fMRI acquisition and analysis*

Participants were scanned in a 3T Philips MRI scanner at Rikshospitalet, Oslo. Each scanning session started with an anatomical scan (0.5 mm^3). Four runs of functional images were acquired while subjects performed the MOT task. One run consisted of 24 trials in two blocks of 12 trials. The load and crowding conditions were semi-randomized. A rest period of 20 s always followed after each block of trials. Whole-brain functional images were acquired using a spin-echo echo-planar (EPI) sequence sensitive to blood-oxygen-level-dependent (BOLD) magnetic susceptibility (TR = 2208 ms; Flip angle = 90deg; number of slices: 42; voxel size: 3 mm^3). Each functional scan lasted about 8 minutes and in each scan, 234 volumes were collected. Finally, a neuromelanin-sensitive scan was acquired with the objective of further improving the localization of the locus coeruleus. The slices for this scan were set in order to cover the brainstem (T1-TSE, TR: 600ms, TE: 14ms, voxel dimension: 0.4x0.49x3mm, Flip angle = 90deg; number of slices: 10). This scan lasted 12 minutes.

The stimuli were projected onto a screen positioned at the head end of the litter. Participants

viewed the screen through a mirror placed on the head coil. Participants' response was given with the right hand through a joystick, with one button corresponding to "target" and another to "not a target" response.

The functional images of each participant were first visually inspected for anomalies and then submitted to a standard preprocessing pipeline using SPM 12 implemented on MATLAB (MathWorks, Natick, MA). Images were first corrected for time delays and realigned using 6 parameters of movement. The data were normalized to a standard template and smoothed (8mm FWHM Gaussian kernel). For the analysis of the activity in the locus coeruleus, a smaller Gaussian kernel was used (3mm FWHM). Additional motion correction was performed by scrubbing the volumes with excessive movement using FSL functions.

Preprocessed images were submitted to a first level analysis. Event-related activation was estimated with a general linear model (GLM). Stimulus presentation and tracking intervals of the different load and crowding levels were modeled as separate events with a canonical HRF. Images were high-pass filtered at 128 sec. Contrasts were generated for each tracking condition both collapsing across load levels and for each load level. Parameter estimates from each participant's GLM were submitted to a second-level test. A full-factorial analysis was performed treating participants as a random factor, and load and crowding as the fixed factors. The maps of load 4 > load 2 and high > low crowding were masked by the respective (load or crowding) main effects. To obtain distinct areas in load and crowding, each of the former maps was masked by the other using thresholds of $p = 0.05$ and $p = 0.001$ for the exclusive mask and the contrast, respectively.

Given our interest in the brainstem, we separately assessed the activity in the VTA/SN and LC. For the VTA/SN, we performed a small volume correction in the contrasts of interest on a combined VTA/SN mask derived from probabilistic atlases of the structures (41). The VTA and SN masks partially overlap and therefore we combined them; the resulting mask was cropped to be spatially restricted to the brainstem. For the LC, given its small size and the variability of its location in the brainstem across individuals, we defined individual masks in the high-resolution T1-TSE images and used them for a region of interest analysis. Specifically, the LC of each individual was delineated on the axial slices of the neuromelanin scans. The position of the nuclei was determined in the pons as the voxels of hyperintensity on either side of the fourth ventricle, following the procedure employed in previous studies (42). The T1-TSE scan of each individual was coregistered to the corresponding structural image, the structural was coregistered to the mean functional and then the deformation field calculated during the normalization step was applied to the coregistered T1-TSE image. Region of interest (ROI) analyses were performed on the individual masks of the locus coeruleus. The data were extracted from the contrast images with rfxplot toolbox (43). The parameter estimates in the voxel of maximum effect size in each condition were extracted for the different load levels within the masks. Then we performed RM-ANOVA applying load (levels: 2, 3 and 4 objects) and crowding (low, high) as within-subject factors. For the comparison of groups, we included group (high-performers and low-performers) as between-subjects factor. For the comparison of task-related activity in the different groups in the VTA, a similar ROI analysis was applied on the VTA/SN combined mask.

2.4 *Data analysis: behavior*

Errors, reaction times, confidence estimates, pupil change, number of saccades and brainstem activity for each condition were compared with repeated measures (RM) ANOVA, correcting for non-sphericity. Within-subject factors were load (3 levels) and crowding (2 levels). Only correct trials were included in the RT analysis. For the analysis of the groups of performers, we divided the participants into high and low performers on the basis of a median split of total tracking accuracy. We performed RM-ANOVA as before adding group as a between subject factor. For this analysis, only the lowest and highest load levels were considered as to look at the largest effects. All effect

sizes are reported as generalized eta squared (η_G^2). The data was analyzed with IBM SPSS.

3. Results

3.1 *Eye-tracking study*

3.1.1 *Behavior*

Increasing load caused a moderate decrease in accuracy ($F(1.92,76.67) = 20.92$, $p < 0.001$, $\eta_G^2 = 0.07$, Figure 2). Furthermore, high crowding also caused a decrease in accuracy ($F(1,40) = 510.56$, $p < 0.001$, $\eta_G^2 = 0.54$), and interacted weakly with load ($F(1.68,67.22) = 5.99$, $p = 0.006$, $\eta_G^2 = 0.01$). Confidence estimates were calculated for each trial as the percentage of responses that were given with the left mouse button (high confidence). An RM-ANOVA on confidence estimates showed significant main effects of load ($F(1.39, 55.52) = 25.27$, $p < .0001$, $\eta_G^2 = 07$) and crowding ($F(1, 40) = 50.64$, $p < .0001$, $\eta_G^2 = .18$), with lower confidence at higher difficulty in both effects. The interaction was not significant ($F(1.69, 67.51) = 3.29$, $p = 0.051$, $\eta_G^2 = .003$).

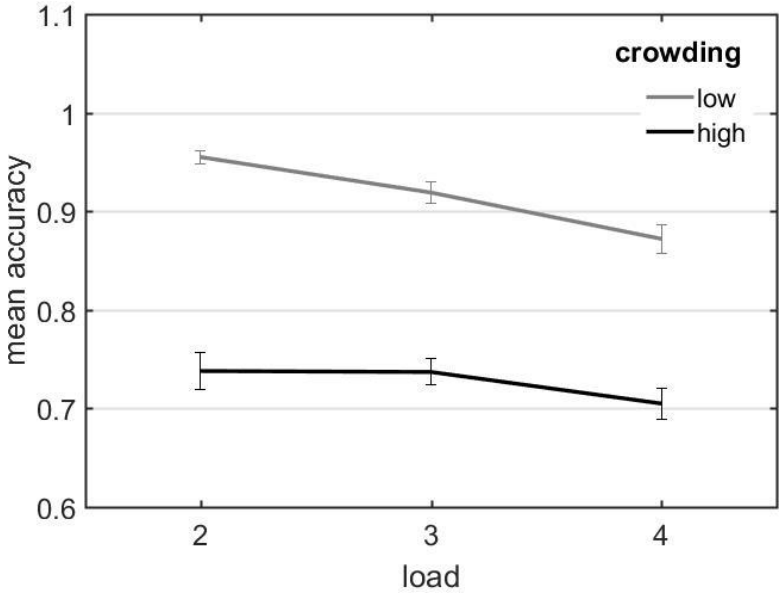

Figure 2 Behavioral results of pupillometry study. Mean accuracy as a function of load for each of the crowding levels.

3.1.2 *Pupil response*

We analyzed the effects of load and crowding on pupil dilation during tracking. We found a significant increase in pupil size with load (main effect of load: $F(1.77,70.86) = 57.6$, $p < 0.001$, $\eta_G^2 = 0.04$; Figure 3). Crowding ($F(1,40) = 0.08$, $p = 0.78$), or the interaction between load and crowding ($F(1.96, 78.25) = 1.26$, $p = 0.29$), did not show significant effects.

When analyzing the pupil change early or late within the trial, the effect of load was significant in both intervals (Early: $F(1.93, 77.25) = 48.73$, $p < 0.001$, $\eta_G^2 = 0.04$; Late: $F(1.83, 73.37) = 34.27$, $p < 0.001$, $\eta_G^2 = 0.02$; Figure 3). The effect of crowding or interaction was non-significant in any interval ($p > 0.28$).

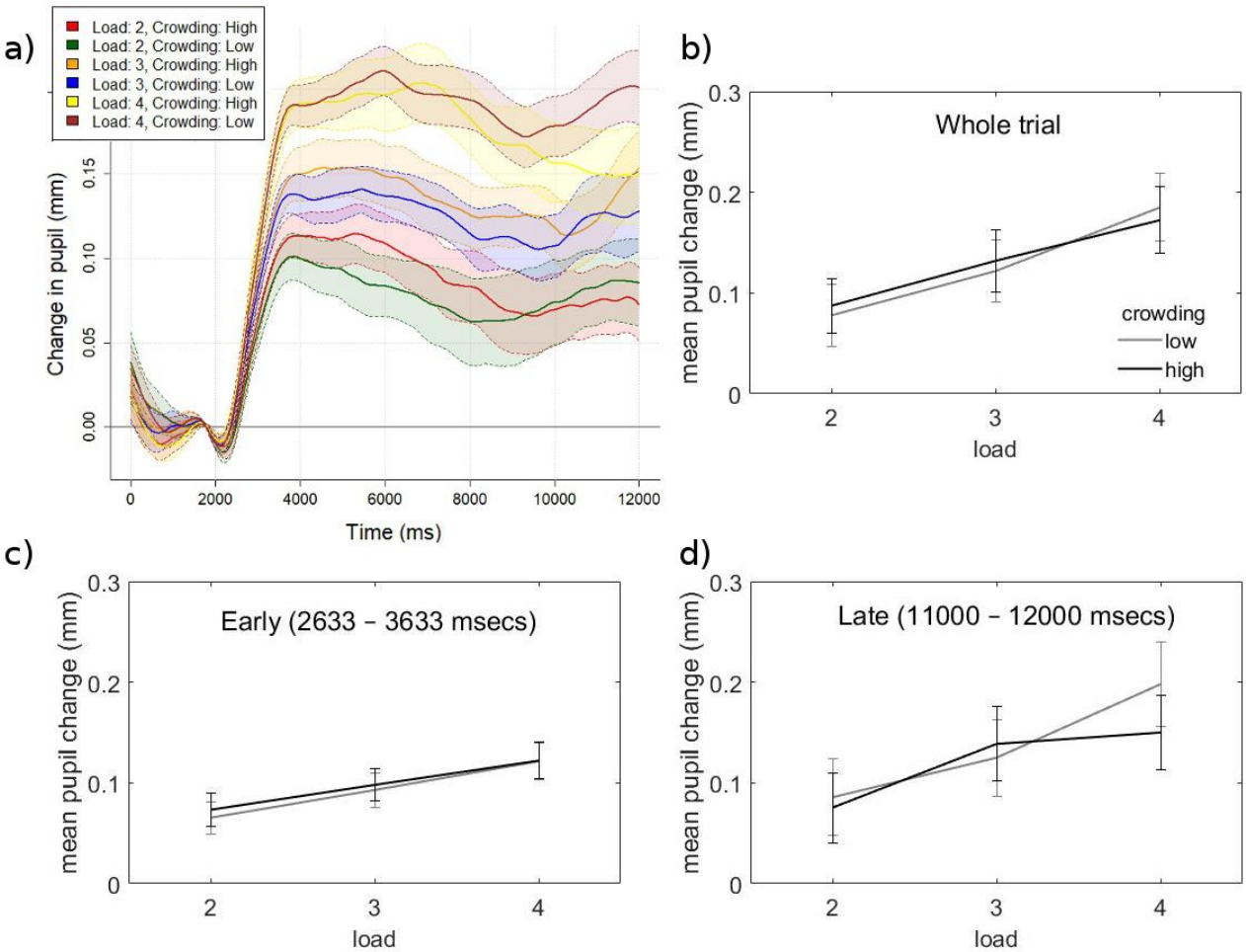

Figure 3. Evoked pupil size during tracking for the different conditions. a) Baseline corrected change in pupil during tracking at the indicated conditions. b-d) Average pupil changes across the whole trial (a), an early interval (b) or a late interval (c) within a trial. Error bars indicate the standard error of the mean.

3.1.3 *Saccade frequencies*

For this analysis, we removed two participants for having an unusual high number of saccades (more than an average of 25 saccades per trial). Number of saccades showed a significant effect of load ($F(1.59, 60.56) = 7.01$, $p = 0.004$, η_G^2 0.004), and crowding ($F(1, 38) = 34.88$, $p < 0.001$, $\eta_G^2 = .02$). The interaction was also significant ($F(2.00, 75.95) = 3.18$, $p = 0.047$, $\eta_G^2 = 0.0006$). On average, participants made 11.1 saccades in high crowding trials and 9.2 saccades in low crowding trials. If participants performed so-called "rescue saccades" (44) toward targets we should observe saccades traveling relatively far from the center fixation dot. Thus, we conducted an ANOVA on number of saccades that ended further than 2 degrees from the center. This revealed main effects of load ($F(1.37, 52.22) = 8.07$, $p = .003$, $\eta_G^2 = .006$) and crowding ($F(1, 38) = 27.12$, $p < .001$, $\eta_G^2 = .02$). The interaction was also significant, ($F(1.88, 71.35) = 4.96$, $p = .01$, $\eta_G^2 = .0009$).

3.1.4 *Individual differences*

We inspected the differences in behavior and ocular measures between high and low performers due to load and crowding (Figure 4). There was a significant interaction between group and load ($F(1.95, 75.94) = 3.85$, $p = 0.030$, $\eta_G^2 = 0.03$) and group and difficulty ($F(1,39) = 27.65$, $p < 0.001$, $\eta_G^2 = 0.08$) and three-way interaction ($F(1.85, 72.2) = 10.69$, $p < 0.001$, $\eta_G^2 = 0.04$).

The analysis of group differences on evoked pupil revealed a significant effect of load ($F(1.77, 69.18) = 56.24$, $p < 0.001$, $\eta_G^2 = 0.04$), and an interaction between performer group, load and difficulty ($F(1.99, 77.45) = 5.17$, $p = 0.008$, $\eta_G^2 = .003$). Low performers had significantly less pupil dilation in high crowding compared to low crowding in load 4 ($t(19) = 3.31$, $p = 0.003$), but not high

performers (p = 0.31).

Next, we analyzed saccade frequencies between high and low performers. High performers made an average of 8.7 saccades per trial, while low performers made an average of 11.9 saccades per trial, this difference was however not statistically significant (t(25.9) = 1.5, p = 0.14). Both of the removed participants were classified as low performers.

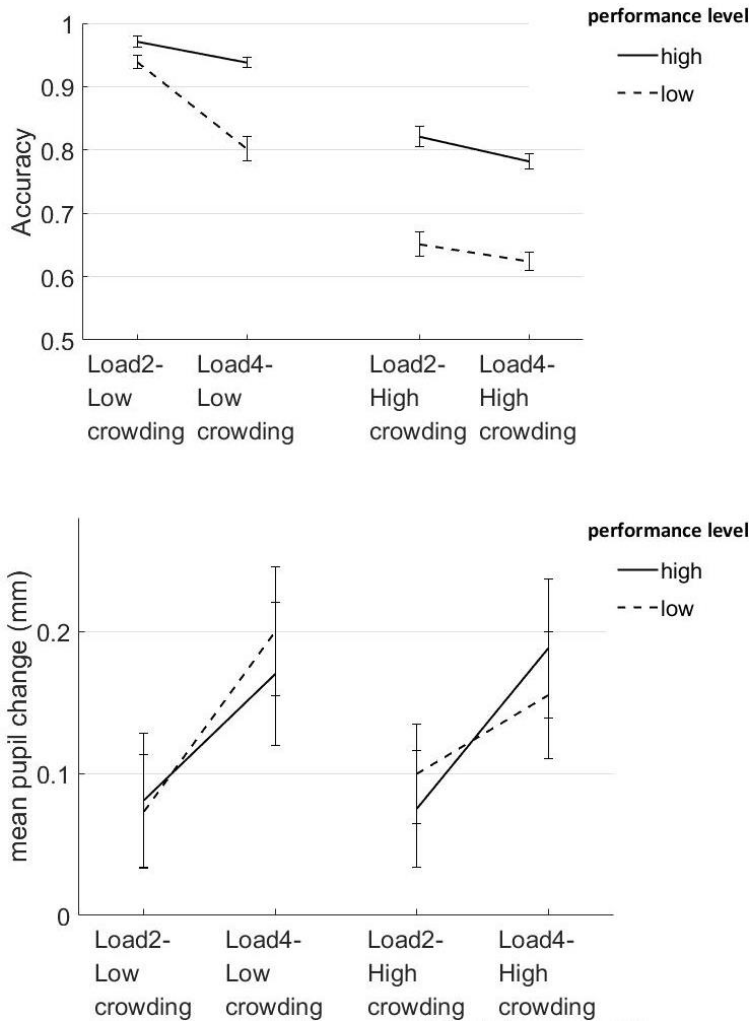

Figure 4 Behavioral and pupillometry results. Mean accuracy (top panel) and mean pupil response (bottom panel) as a function of load for each of the crowding levels.

3.2 *fMRI study*

3.2.1 *Behavior*

Increasing load caused a moderate decrease in accuracy ($F(2,78) = 12.26$, $p < 0.001$, $\eta_G^2 = 0.08$, Figure 5). Furthermore, high crowding caused a large decrease in accuracy ($F(1,78) = 120.23$, $p < 0.001$, $\eta_G^2 = 0.25$), and did not interact with load ($F(2,78) = 0.40$, $p = 0.63$). RT increased with higher load ($F(2,78) = 4.65$, $p = 0.012$, $\eta_G^2 = 0.01$) and higher crowding ($F(1,78) = 27.71$, $p < 0.001$, $\eta_G^2 = 0.02$). There was no interaction between load and crowding in RTs ($F(2,78) = 0.74$, $p = 0.45$).

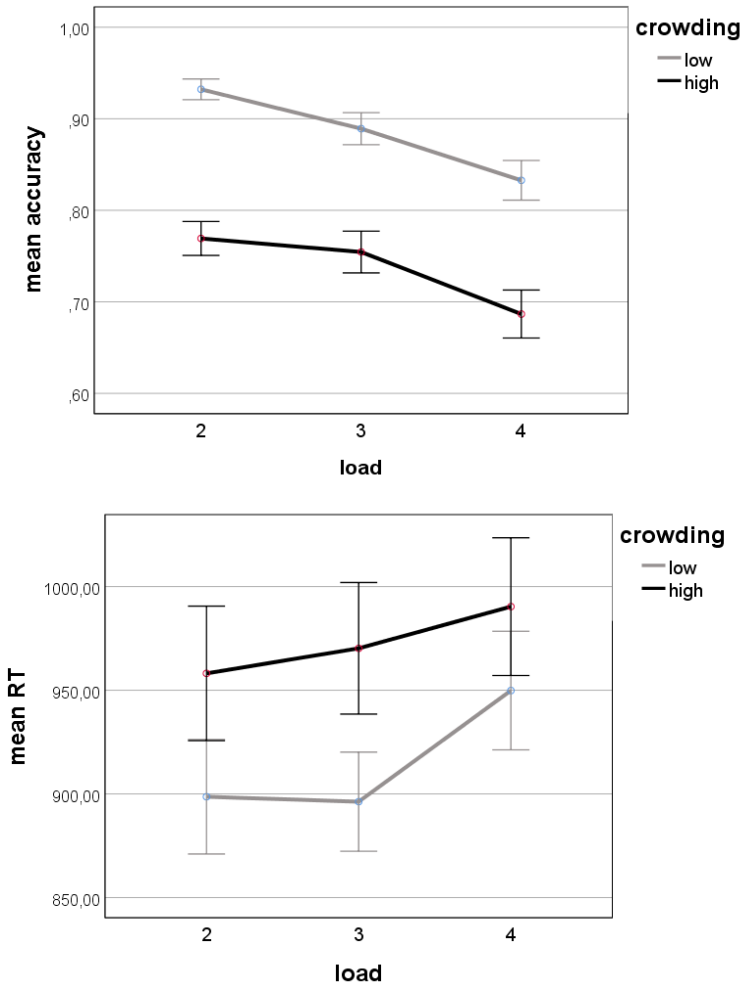

Figure 5 Behavioral results of fMRI study. Mean accuracy (top panel) and RT (bottom panel) as a function of load and for each crowding level. Error bars indicate the standard error of the mean.

3.2.2 *Whole brain fMRI analysis*

A set of areas was associated to tracking as compared to passive viewing, comprising the bilateral inferior/middle occipital cortex, the bilateral inferior/superior parietal cortex and bilateral middle frontal cortex, precentral area and supplementary motor area. (Figure 6, Tables 1 and 2). This pattern replicates previous findings on MOT and fMRI (15).

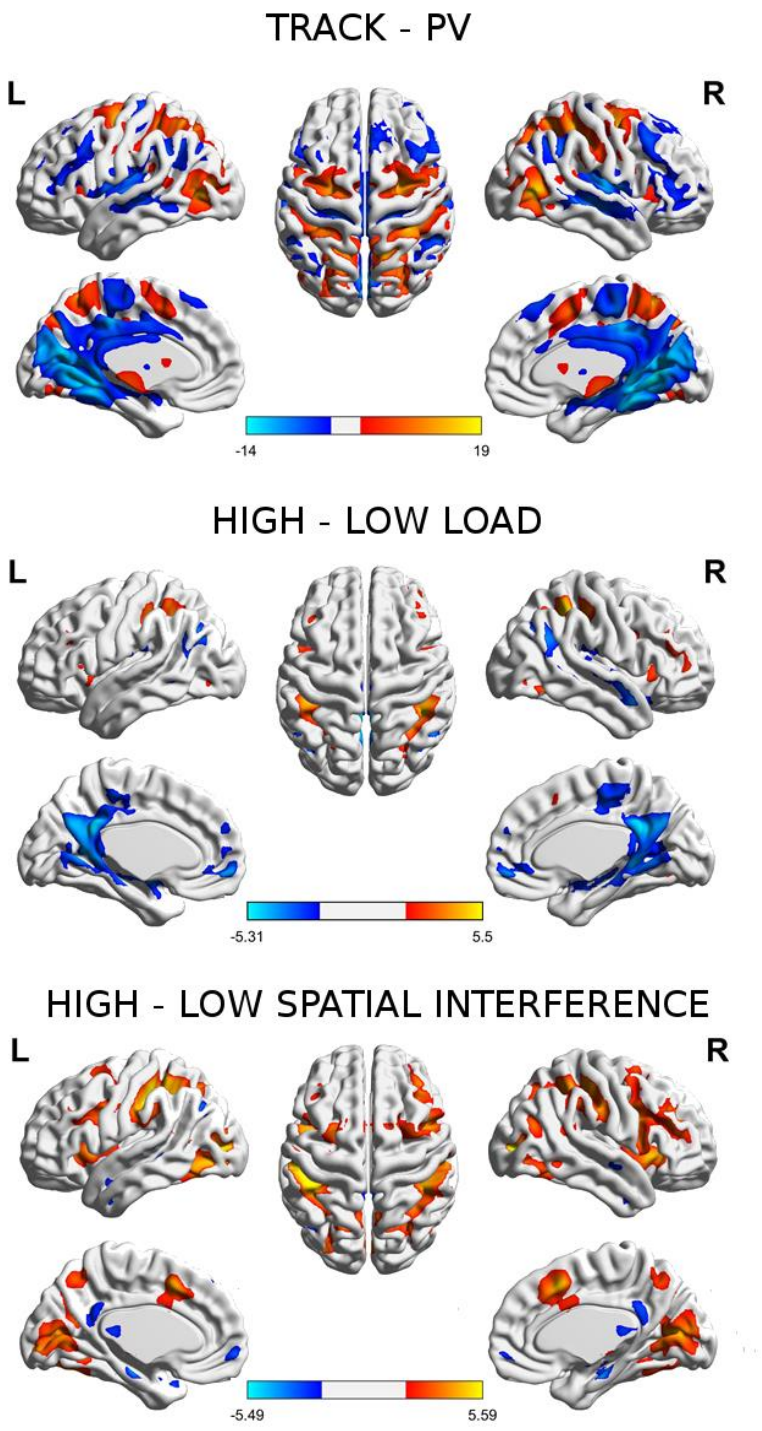

Figure 6 Brain regions' activations and deactivations obtained from the indicated contrasts. L: left hemisphere; R: right hemisphere. The colorbar indicates the obtained t value.

Load evoked parametrically higher activation in bilateral inferior parietal area and left inferior frontal (Figure 6, Tables 1 and 3). High crowding evoked higher activity in middle occipital areas, as well as bilateral inferior parietal and bilateral middle frontal gyrus/precentral area including frontal eye fields, and insula (Figure 6, Tables 1 and 3). The difference between these contrasts revealed that crowding presented additional activity in bilateral inferior parietal and medial frontal regions, insula and primary visual areas (lingual/calcarine) as compared to load (Figure 7, Table 3).

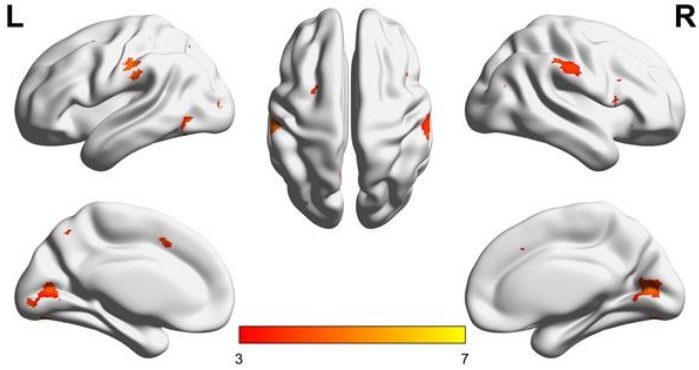

Figure 7 Effect of high – low crowding (masked Load 4 – load 2)

3.2.3 *Brainstem nuclei*

We inspected the effect of increasing effort on the activity within individually defined LC masks. We found a significant effect of load ($F(2,74) = 3.21$, $p = 0.046$, $\eta_G^2 = 0.02$, Figure 8). We found no effect of crowding ($F(1,37) = 0.59$, $p = 0.44$) or interaction ($F(2,74) = 3.21$, $p = 0.066$).

Moreover, load had no significant effect on activity in the VTA/SN nuclei ($F(2,74) = 1.71$, $p = 0.194$, Figure 8) but high crowding evoked a larger activity than low ($F(1,37) = 4.96$, $p = 0.032$, $\eta_G^2 = 0.02$). We found no interaction ($F(2,74) = 0.375$, $p = 0.656$).

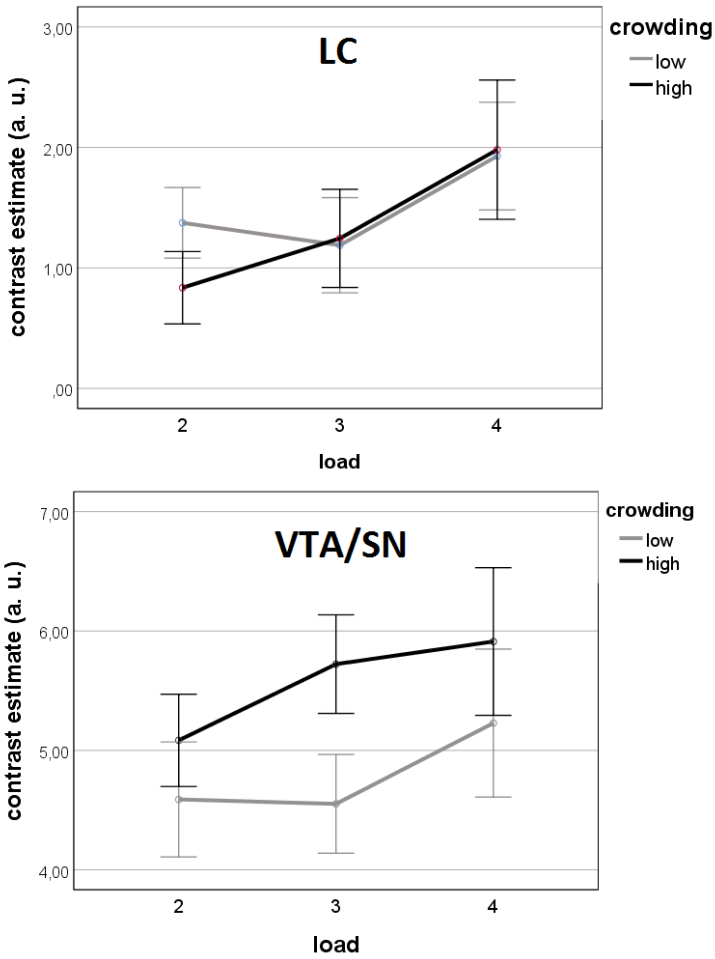

Figure 8 Brainstem activity evoked by the different conditions. Top: LC; bottom: VTA/SN

3.2.4 *Individual differences*

High and low performers showed different levels of accuracy for the different conditions. In addition to the main effects of load and crowding ($F(1,36) = 28.51$, $p < 0.001$ $\eta_G^2 = 0.11$ and $F(1,36)$

= 88.14, $p < 0.001$, $\eta_G^2 = 0.27$, resp., Figure 9), we found a significant interaction of performance level by load ($F(1,36) = 9.20$, $p = 0.004$, $\eta_G^2 = 0.04$). Low performers had lower accuracy in load 4 as compared to load 2 ($T(19) = 5.50$, $p < 0.001$), while the effect of load in high performers did not reach significance ($T(17) = 1.83$, $p = 0.084$). No group effects were observed for RT.

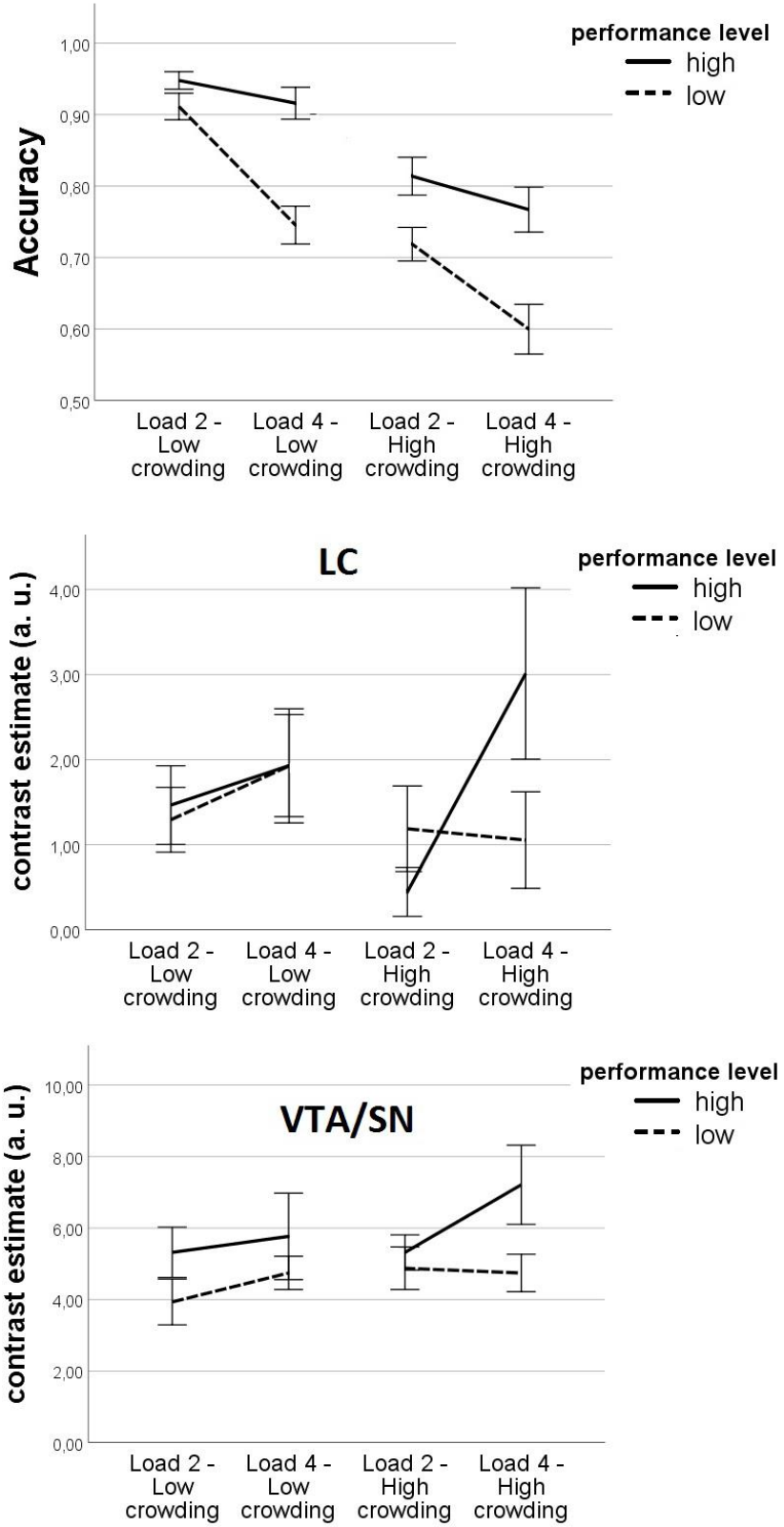

Figure 9 Behavioral effects on high and low performers. Mean accuracy in the high and low crowding conditions for the lowest and highest load level split by high performers versus low performers.

In the LC, we found a significant three-way interaction of load by performance level by crowding ($F(1,36) = 4.44$, $p = 0.042$, $\eta_G^2 = 0.02$, Figure 9). When looking separately at the different crowding levels, we observed a significant load by group interaction ($F(1,36) = 5.55$, $p = 0.024$, $\eta_G^2 = 0.06$) in the high but not in the low crowding condition. With high crowding, high performers had higher activity in load 4 as compared to load 2 ($T(17) = 2.51$, $p = 0.022$), but not low performers ($p > 0.2$ for all contrasts).

In the VTA/SN, we found no main effect or interaction effects (Figure 9). High performers presented larger activity than low performers, although the results were not significant ($F(1,36) = 3.17$, $p = 0.083$).

4. Discussion

Based on the present findings, we conclude that task load and spatial interference are handled by distinct processes that rely, at least partially, on different neural mechanisms. This conclusion is based on the observation that, even when the level of spatial interference (low/high crowding) was kept equal, accuracy decreased and RT increased with load level (number of targets), hinting that crowding and load pose challenges to brain processing independently. In line with this, increasing demands due to load were signaled by increasing arousal (as indicated by pupil size) and LC activity across the tracking period. Increased effort recruited proportionally greater activation in the fronto-parietal attention network and saccades. In addition, increasing crowding events engaged the VTA/SN, and further implicated the attentional network and number of saccades. Furthermore, high performers appeared able to recruit their cognitive resources during stimulus-driven conflict, as signaled by higher LC activity during high crowding. Taken together, the above results suggest that the brain can cope with spatial interference and effort in a highly specific and adaptive manner.

4.1 *Behavioral effects of load and spatial interference*

We note that different factors are known to affect performance in MOT: number of objects or cognitive workload (both targets and non-targets), crowding, speed and hemifield (2, 3, 45). All these factors seem to share a common property: they affect the degree and frequency of proximity between objects. Hence, a still open question regarding the MOT paradigm was whether the deterioration of performance could be simply reduced to spatial interference, that is, to a perceptual limitation in attentional resolution. For example, increasing number of targets could imply that more close encounters with other objects will likely occur and, therefore, load should correlate with number of close encounters. Some studies have assessed the effect of inter-object spacing on accuracy, either at different or at a fixed level of load (7-12). Others have assessed the relation between load and speed (13, 46) or between speed and object spacing (10, 47). In the current study, we systematically varied load at two levels of crowding. This allowed us to assess the relation between the factors.

A decreasing linear trend in accuracy with load at both levels of crowding suggests that, in addition to the frequency of close encounters, the number of objects pose an additional demand on the participants, and therefore those factors rely on partly non-overlapping mechanisms (Figure 1.b). Given that the spatial interference was equalized within crowding levels and across load demands, our results indicate that load may pose an independent demand on attentional functions. This agrees with previous work suggesting that recruiting higher order mechanisms allow for dynamic allocation of attention to crowded situations (9, 14, 46, 48). Such higher order mechanisms likely involve visual selection and sustained visual attention, which may be specifically engaged during target assignment or during the tracking period, respectively (3). While the above terms may seem

somehow loosely defined mechanistically, it has been shown that cognitive processes, such as update in working memory, are engaged at the assignment of objects (49, 50) and during tracking (6, 51), consistent with computational simulations (46, 47). Similarly, Vul, Alvarez (47) argued that "while many results in MOT arise as consequences of the information available for the computational task, the speed-number tradeoff seems to be the result of a flexibly-allocated resource such as memory or attention". Possibly, cognitive processes may help to circumvent the limitations of the visual attentional processes and viceversa (6). In line with this, a series of EEG studies have shown that different ERP signatures were related to the demands exerted by target number and the demands of the tracking itself (52-54). For example, a parieto-central negativity was related to increasing target number during the assignment period, a contralateral delay activity was related to load during the tracking period, and early negative and positive components were related to following targets rather than distractors during the tracking period. These results argue for a highly coordinated and dynamic process developing during the tracking to keep target information in working memory, resolve confusion and remain engaged.

4.2 *Pupil increases with load but not with crowding*

We replicated the finding that pupil size is sensitive to load (17, 40). One possibility is that pupil size reflects preparatory effort mechanisms recruited in the beginning of the trial and sustained during the whole trial. This would agree with the proposal that pupil dilation reflects a readiness to expend resources (32). Importantly, we found that pupil size was not sensitive to our manipulation of spatial interference. Thus, one could argue that the effects of load and crowding rely on different mechanisms, for example different neural mechanisms (as discussed below). We also note that effort due to the load manipulation is sustained over the tracking period, whereas the close encounters are transient events. It was expected that the effect of such transient events would accumulate to such an extent that they would influence the average pupil size over the entire tracking period, but it is also possible that these effects would wash out in the average.

The number of saccades increased with load and crowding, although in a greater extent with crowding. This result may correspond to the increase in activity that we observed in the frontal eye fields (FEF) in high versus low crowding. FEF are involved in the control of eye movement (55) and may have being implicated in directing the eyes towards highly conflicting events (i.e. close encounters), what is consistent with previous results on MOT and eye-tracking (44).

4.3 *Cortical attentional networks related to both load and spatial interference*

We found that a fronto-parietal network including inferior parietal and dorsal frontal areas, as well as occipital areas, were recruited during tracking. Load was further related to increased activity in the bilateral inferior parietal and inferior frontal areas. These activations closely match the ones reported by previous studies on MOT (15-21). The fronto-parietal network is generally activated during demanding visuo-spatial tasks of different nature (22, 24, 56), suggesting that its components cooperate to represent voluntary and stimulus-driven attentional priority (23, 57, 58). As part of this network, the inferior parietal cortex is suggested to index the location of selected or prioritized objects (57) and its activity depends on capacity limitations in visual short-term memory (59). The inferior parietal cortices interact with superior parietal and frontal areas (60) and constitute the so-called dorsal attention network (24, 61, 62), whose involvement is consistent with top-down processes engaged during the task to guide the tracking of the target objects.

Higher crowding presented strong activity in bilateral inferior parietal, dorsal-frontal, including FEF, and occipital regions. In addition, insula and medial frontal areas were engaged. The insula and medial frontal areas are considered parts of the ventral attention network (62, 63), whose function is thought to be detecting salient events and assist attention orienting. During tracking, events

such as close encounters may engage the ventral attention network to flexibly allocate attention where it is most needed. Therefore, an increased activity in high vs. low crowding may be related to coping with the increasing on-line demands imposed by that manipulation.

The frontal eye fields were found to be activated during high versus low crowding. In a seminal study on MOT with fMRI, Culhan et al. (15) found that the activity in FEF was related to tracking but was not modulated by load. Here we corroborate this finding, and in addition we propose that the FEF activity during tracking is related to the crowding events, which were not contolled for in Culhan et al. study. Therefore, its involvement in the task seems to be active in the resolution of moment-to-moment confusion, rather than barely "suppressing eye movements" (15).

Previous studies showed specificity in the factors accounting for the activation of some areas during multiple-object tracking (60, 64, 65). Here we show that large activations in attentional and occipital areas were observed due to spatial interference as compared to load. Strong activation in the medial superior parietal lobe (including precuneus) has been suggested to reflect voluntary shifts of attention between perceptual entities (57, 66-68). Other areas, such as medial frontal cortex and insula, were recruited additionally during high vs. low crowding and as compared to the load effect. A note of caution should be made given that it is not possible to know whether load 4, compared to load 2, imposed as high demands as high vs. low crowding, as indicated by the different effect sizes of those two factors on accuracy. Indeed, previous studies have shown larger brain effects of load when including load 5 (e.g. (17)). Therefore, we cannot exclude that the network of cortical areas underlying load and crowding were the same and that differences in the demand levels of our paradigm caused that the effects of crowding in brain activity to be stronger than the effect of load.

4.4 *Brainstem neuromodulatory systems*

Neuromodulators play an active role on attention by adjusting responsivity of neurons to optimize processing of inputs, a concept known as "gain" (69, 70). In the context of the MOT, we found that two nuclei, locus coeruleus (noradrenergic) and VTA/SN (dopaminergic), presented different profiles for the different task manipulations. LC increased activity with load, while VTA/SN showed increased activity with spatial interference. The involvement of LC agrees with the findings from a previous study which employed an atlas-based mask of the LC to extract the task-related activity (17). The LC/NE system has been studied in the context of mental effort and its function has been posited to be to increase the gain of the neural circuits relating co-occurring events. This result may be explained by the recruitment of LC by a top-down frontal system engaged in the preparation of load-related effort. We further found that LC activity was not significantly related to spatial interference. In contrast, the VTA/SN presented higher activity in high with respect to low spatial interference, but no effect with cognitive workload. Such a double dissociation between brainstem systems is (to the best of our knowledge) novel within human imaging studies in the context of an attentional task with varying levels of cognitive effort. Interestingly, the VTA/SN and LC share reciprocal connections, and therefore they are expected to be tightly related. Within theories of cognitive effort, LC activity may be recruited after the computation of cost at the initiation of the trial, when the targets' number or load is known. VTA/SN activity may be, on the other hand, related to working memory processes that enable propagation of state estimates of the position of the targets through time. In other words, one mechanism (VTA/SN) may depend from events occurring in spatiotopic regions, whereas the other (LC) may be essentially non-spatial and more related to temporal relations between events and the recruitment of resources (effort or arousal). While coherent with the roles associated to these neuromodulatory systems, the ventral attentional network (which we found to be involved during spatial interference), has been

proposed to recruit LC after detection of salient events (63); however, since we did not observe that LC was related to changes in spatial interference, its role may mainly relate to non-spatial aspects of the attention network.

4.5 *Individual differences*

The analysis of individual differences showed that, across studies, low performers' accuracy decreased with load compared to high performers. The effect of crowding was significant in the pupillometry but not in the fMRI study. This result suggests that the processes of load and crowding are independent in terms of compromising participants' performance, something that has been already proposed (6, 14). The pupil results showed that high performers showed pupil dilation with load both in low and high crowding, suggesting that they were able to allocate resources even in the most difficult condition. The brainstem analysis on individual differences indicated that, in addition to being recruited with increasing load, the LC was also to some extent recruited with increasing crowding but only in high performing subjects. Together, these results seem to reflect these participants' ability to flexibly recruit resources to keep high levels of accuracy in a highly challenging condition. Thus, our findings suggest the intriguing notion that high performers are not only better prepared to resolving high conflict (i.e. high load) trials, but also able to recruit most effectively their cognitive resources during cognitive conflict.

Together with the cholinergic systems, the dopaminergic and noradrenergic systems modulate and facilitate visuo-spatial attention and visual working memory processes (71, 72). Individual differences are related to genetic and physiological variations in these systems, and they may underlie variations in aging as well as states of mental illnesses (39, 71, 73).

4.5 *Relation with theories on limitations in dynamic visual attention*

Theoretically, two competing models have been proposed to account for the perceptual and cognitive limitations observed during a task requiring dynamic attention, such as MOT (45, 74). One is the coding of individual stimuli into cortical spatial maps that may inhibit each other due to close perceptual proximity of the stimuli (cortical real-estate account). Another is the coding of stimuli in separate working memory slots ("resource" account). While the first predicts that load and crowding engage the same substrates in primary visual circuits, the second would predict that different networks deal with load and crowding (that is, objects are represented as informational chunks and additional mechanisms that deal with spatial interference due to crowding are recruited as needed). Although our results do support the cortical real-state account, given that spatial interference influences accuracy and cortical as well as some of the brainstem activations, this view encounters difficulty in explaining effects on accuracy, cortical and brainstem activation, and pupil dilations when spatial interference is held constant. Moreover, we observed activations in LC and pupil responses related to load and not to the spatial aspects of crowding. Hence, a general "resource" account as originally proposed by Kahneman (28), although it may be unsatisfactory for a number of reasons (45), remains a candidate explanation. However, we note that the above two views have mainly focused on the perceptual level. Previous studies suggest that not only perceptual but also executive processing is crucial for resolving MOT, not only during tracking but also at the moment of target selection (4, 47, 49-51, 75, 76). Therefore, a resolution between the two hypotheses is not straightforward based on the present results. Nevertheless, MOT shares many characteristics with other visuo-spatial tasks that rely on perceptual as well as executive processes (i.e. VSTM, (77), (78); subitizing: (79); selection of objects: (80)) and consequently, its study may inform how limitations in the other processes occur.

4.6 *Study limitations*

For the present work, we only had two levels of crowding (i.e. spatial interference). Future

experiments with more levels of crowding in addition to load would allow a more complete picture of the relation between the two factors. Although we found no interaction between them, adding crowding levels may present a more complex picture. Another remaining question is what would be the effect of informing participants at the start of the trial whether there will be high or low crowding in the display at the same time as telling the number of targets. A prediction would be that, if preparatory activity related to load would be engaged to prepare for higher crowding, the accuracy patterns of load and crowding would be dependent (Figure 1.c). For the purpose of this study, we focused on the role of attention on resolving spatial interference during tracking and therefore did not study brain and brainstem activity during visual selection (i.e. activity during the target assignment phase). Studies focusing on this stage may confirm previous results suggesting that executive functions are involved in this stage and influence tracking, and extend our results regarding the involvement of brain and brainstem regions in the encoding of targets.

5. Conclusion

Our results support the view that, in the MOT task, distinct mechanisms deal with the increasing demands posed by cognitive workload (i.e. number of objects) and spatial interference (i.e. crowding), and brainstem neuromodulatory nuclei play a key role. Load, independently from spatial interference, compromised accuracy during MOT. A set of regions within the dorsal attention network were recruited with both load and spatial interference demands. Additional areas within the ventral attention network were recruited during high crowding. The ventral and dorsal attentional systems collaborate to solve the challenges of covert goals, likely engaging top-down control, as well as rising demands, likely engaging bottom-up processes (62, 63). Locus coeruleus and pupil dilation signaled effort in the sense of engagement, which occurred with increasing load and, in high performers, with higher crowding. VTA/SN differed from LC in that it was related to spatial interference. The MOT task, together with physiological measures, may offer a unique opportunity to interrogate how cortical and neuromodulatory systems are adaptively recruited when cognition is challenged.

References

1. Pylyshyn ZW, Storm RW. Tracking multiple independent targets: Evidence for a parallel tracking mechanism. Spatial vision. 1988;3(3):179-97.
2. Scimeca JM, Franconeri SL. Selecting and tracking multiple objects. Wiley Interdisciplinary Reviews: Cognitive Science. 2015;6(2):109-18.
3. Meyerhoff HS, Papenmeier F, Huff M. Studying visual attention using the multiple object tracking paradigm: A tutorial review. Attention, Perception, & Psychophysics. 2017;79(5):1255-74.
4. Tombu M, Seiffert AE. Attentional costs in multiple-object tracking. Cognition. 2008;108(1):1-25.
5. Meyerhoff HS, Papenmeier F, Jahn G, Huff MJJoEPHP, Performance. Not FLEXible enough: Exploring the temporal dynamics of attentional reallocations with the multiple object tracking paradigm. 2016;42(6):776.
6. Oksama L, Hyönä J. Is multiple object tracking carried out automatically by an early vision mechanism independent of higher-order cognition? An individual difference approach. Visual cognition. 2004;11(5):631-71.
7. Bae GY, Flombaum JI. Close encounters of the distracting kind: Identifying the cause of visual tracking errors. Attention, Perception, & Psychophysics. 2012;74(4):703-15.
8. Feria CS. Speed has an effect on multiple-object tracking independently of the number of close encounters between targets and distractors. Attention, Perception, & Psychophysics. 2013;75(1):53-67.
9. Iordanescu L, Grabowecky M, Suzuki S. Demand-based dynamic distribution of attention

and monitoring of velocities during multiple-object tracking. Journal of vision. 2009;9(4):1-.
10. Alvarez GA, Franconeri SL. How many objects can you track?: Evidence for a resource-limited attentive tracking mechanism. Journal of vision. 2007;7(13):14-.
11. Franconeri SL, Lin JY, Enns JT, Pylyshyn ZW, Fisher B. Evidence against a speed limit in multiple-object tracking. Psychonomic bulletin & review. 2008;15(4):802-8.
12. Shim WM, Alvarez GA, Jiang YV. Spatial separation between targets constrains maintenance of attention on multiple objects. Psychonomic bulletin & review. 2008;15(2):390-7.
13. Franconeri S, Jonathan S, Scimeca J. Tracking multiple objects is limited only by object spacing, not by speed, time, or capacity. Psychological science. 2010;21(7):920-5.
14. Intriligator J, Cavanagh P. The spatial resolution of visual attention. Cognitive psychology. 2001;43(3):171-216.
15. Culham JC, Brandt SA, Cavanagh P, Kanwisher NG, Dale AM, Tootell RB. Cortical fMRI activation produced by attentive tracking of moving targets. Journal of neurophysiology. 1998;80(5):2657-70.
16. Culham JC, Cavanagh P, Kanwisher NG. Attention response functions: characterizing brain areas using fMRI activation during parametric variations of attentional load. Neuron. 2001;32(4):737-45.
17. Alnæs D, Sneve MH, Espeseth T, Endestad T, van de Pavert SHP, Laeng B. Pupil size signals mental effort deployed during multiple object tracking and predicts brain activity in the dorsal attention network and the locus coeruleus. Journal of vision. 2014;14:1-.
18. Shim WM, Alvarez GA, Vickery T, Jiang Y. The number of attentional foci and their precision are dissociated in the posterior parietal cortex. Cerebral Cortex. 2009;20(6):1341-9.
19. Jovicich J, Peters RJ, Koch C, Braun J, Chang L, Ernst T. Brain areas specific for attentional load in a motion-tracking task. Journal of cognitive neuroscience. 2001;13(8):1048-58.
20. Jahn G, Wendt J, Lotze M, Papenmeier F, Huff M. Brain activation during spatial updating and attentive tracking of moving targets. Brain and cognition. 2012;78(2):105-13.
21. Tomasi D, Ernst T, Caparelli E, Chang L. Practice-induced changes of brain function during visual attention: a parametric fMRI study at 4 Tesla. Neuroimage. 2004;23(4):1414-21.
22. Duncan JJTics. The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. 2010;14(4):172-9.
23. Ptak RJTN. The frontoparietal attention network of the human brain: action, saliency, and a priority map of the environment. 2012;18(5):502-15.
24. Corbetta M, Shulman GL. Control of goal-directed and stimulus-driven attention in the brain. Nature reviews neuroscience. 2002;3(3):201.
25. Thiele A, Bellgrove MAJN. Neuromodulation of Attention. 2018;97(4):769-85.
26. Williges RC, Wierwille WW. Behavioral measures of aircrew mental workload. Human Factors. 1979;21(5):549-74.
27. Strobach T, Wendt M, Janczyk M. Multitasking: Executive Functioning in Dual-Task and Task Switching Situations. Frontiers in Psychology. 2018;9:108.
28. Kahneman D. Attention and effort: Citeseer; 1973.
29. Beatty J. Task-evoked pupillary responses, processing load, and the structure of processing resources. Psychological bulletin. 1982;91(2):276.
30. van der Wel P, van Steenbergen H. Pupil dilation as an index of effort in cognitive control tasks: A review. Psychonomic bulletin & review. 2018:1-11.
31. Laeng B, Sirois S, Gredebäck G. Pupillometry: a window to the preconscious? Perspectives on psychological science. 2012;7(1):18-27.
32. Bruya B, Tang Y-YJFip. Is attention really effort? Revisiting Daniel Kahneman's influential 1973 book attention and effort. 2018;9.
33. Varazzani C, San-Galli A, Gilardeau S, Bouret SJJoN. Noradrenaline and dopamine neurons in the reward/effort trade-off: a direct electrophysiological comparison in behaving monkeys. 2015;35(20):7866-77.
34. Ang Y-S, Manohar S, Apps MAJFibn. Commentary: noradrenaline and dopamine neurons in

the reward/effort trade-off: a direct electrophysiological comparison in behaving monkeys. 2015;9:310.
35. Knudsen EIJTin. Neural Circuits That Mediate Selective Attention: A Comparative Perspective. 2018.
36. Murphy PR, O'connell RG, O'sullivan M, Robertson IH, Balsters JHJHbm. Pupil diameter covaries with BOLD activity in human locus coeruleus. 2014;35(8):4140-54.
37. Westbrook A, Braver TS. Dopamine does double duty in motivating cognitive effort. Neuron. 2016;89(4):695-710.
38. Unsworth N, Robison MK. Individual differences in the allocation of attention to items in working memory: Evidence from pupillometry. Psychonomic Bulletin & Review. 2015;22(3):757-65.
39. Unsworth N, Robison MK. The importance of arousal for variation in working memory capacity and attention control: A latent variable pupillometry study. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2017;43(12):1962.
40. Wright TJ, Boot WR, Morgan CS. Pupillary response predicts multiple object tracking load, error rate, and conscientiousness, but not inattentional blindness. Acta psychologica. 2013;144(1):6-11.
41. Murty VP, Shermohammed M, Smith DV, Carter RM, Huettel SA, Adcock RAJN. Resting state networks distinguish human ventral tegmental area from substantia nigra. 2014;100:580-9.
42. Krebs RM, Park HR, Bombeke K, Boehler CN. Modulation of locus coeruleus activity by novel oddball stimuli. 2018;12(2):577-84.
43. Gläscher J. Visualization of group inference data in functional neuroimaging. Neuroinformatics. 2009;7(1):73-82.
44. Zelinsky GJ, Todor AJJoV. The role of “rescue saccades” in tracking objects through occlusions. 2010;10(14):29-.
45. Franconeri SL, Alvarez GA, Cavanagh PJTics. Flexible cognitive resources: competitive content maps for attention and memory. 2013;17(3):134-41.
46. Srivastava N, Vul E. Attention Modulates Spatial Precision in Multiple-Object Tracking. Topics in cognitive science. 2016;8(1):335-48.
47. Vul E, Alvarez G, Tenenbaum JB, Black MJ, editors. Explaining human multiple object tracking as resource-constrained approximate inference in a dynamic probabilistic model. Advances in neural information processing systems; 2009.
48. Pylyshyn Z. Some puzzling findings in multiple object tracking: I. Tracking without keeping track of object identities. Visual cognition. 2004;11(7):801-22.
49. Ma Z, Flombaum JI. Off to a bad start: Uncertainty about the number of targets at the onset of multiple object tracking. Journal of experimental psychology: human perception and performance. 2013;39(5):1421.
50. Allen R, Mcgeorge P, Pearson DG, Milne A. Multiple-target tracking: A role for working memory? The Quarterly journal of experimental psychology. 2006;59(6):1101-16.
51. Eayrs J, Lavie N. Establishing individual differences in perceptual capacity. Journal of Experimental Psychology: Human Perception and Performance. 2018.
52. Drew T, Horowitz TS, Vogel EK. Swapping or dropping? Electrophysiological measures of difficulty during multiple object tracking. Cognition. 2013;126(2):213-23.
53. Drew T, McCollough AW, Horowitz TS, Vogel EK. Attentional enhancement during multiple-object tracking. Psychonomic Bulletin & Review. 2009;16(2):411-7.
54. Drew T, Vogel EK. Neural measures of individual differences in selecting and tracking multiple moving objects. Journal of Neuroscience. 2008;28(16):4183-91.
55. Vernet M, Quentin R, Chanes L, Mitsumasu A, Valero-Cabré AJFiin. Frontal eye field, where art thou? Anatomy, function, and non-invasive manipulation of frontal regions involved in eye movements and associated cognitive operations. 2014;8:66.
56. Hugdahl K, Raichle ME, Mitra A, Specht KJFihn. On the existence of a generalized non-specific task-dependent network. 2015;9:430.

57. Serences JT, Yantis S. Spatially selective representations of voluntary and stimulus-driven attentional priority in human occipital, parietal, and frontal cortex. Cerebral cortex. 2006;17(2):284-93.
58. Scolari M, Seidl-Rathkopf KN, Kastner S. Functions of the human frontoparietal attention network: Evidence from neuroimaging. Current opinion in behavioral sciences. 2015;1:32-9.
59. Mitchell DJ, Cusack R. Flexible, capacity-limited activity of posterior parietal cortex in perceptual as well as visual short-term memory tasks. Cerebral Cortex. 2007;18(8):1788-98.
60. Howe PD, Horowitz TS, Morocz IA, Wolfe J, Livingstone MS. Using fMRI to distinguish components of the multiple object tracking task. Journal of Vision. 2009;9(4):10-.
61. Fox MD, Corbetta M, Snyder AZ, Vincent JL, Raichle ME. Spontaneous neuronal activity distinguishes human dorsal and ventral attention systems. Proceedings of the National Academy of Sciences. 2006;103(26):10046-51.
62. Vossel S, Geng JJ, Fink GR. Dorsal and ventral attention systems: distinct neural circuits but collaborative roles. The Neuroscientist. 2014;20(2):150-9.
63. Corbetta M, Patel G, Shulman GL. The reorienting system of the human brain: from environment to theory of mind. Neuron. 2008;58(3):306-24.
64. Merkel C, Hopf J-M, Heinze H-J, Schoenfeld MA. Neural correlates of multiple object tracking strategies. Neuroimage. 2015;118:63-73.
65. Atmaca S, Stadler W, Keitel A, Ott DV, Lepsien J, Prinz W. Prediction processes during multiple object tracking (MOT): involvement of dorsal and ventral premotor cortices. Brain and behavior. 2013;3(6):683-700.
66. Yantis S, Schwarzbach J, Serences JT, Carlson RL, Steinmetz MA, Pekar JJ, et al. Transient neural activity in human parietal cortex during spatial attention shifts. Nature neuroscience. 2002;5(10):995.
67. Serences JT, Shomstein S, Leber AB, Golay X, Egeth HE, Yantis S. Coordination of voluntary and stimulus-driven attentional control in human cortex. Psychological science. 2005;16(2):114-22.
68. Serences JT, Liu T, Yantis S. Parietal mechanisms of attentional control: locations, features, and objects. Neurobiology of attention: Elsevier; 2005. p. 35-41.
69. Servan-Schreiber D, Printz H, Cohen JD. A network model of catecholamine effects: gain, signal-to-noise ratio, and behavior. Science. 1990;249(4971):892-5.
70. Aston-Jones G, Cohen JD. An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. Annu Rev Neurosci. 2005;28:403-50.
71. Störmer VS, Passow S, Biesenack J, Li S-C. Dopaminergic and cholinergic modulations of visual-spatial attention and working memory: insights from molecular genetic research and implications for adult cognitive development. Developmental psychology. 2012;48(3):875.
72. Sarter M, Gehring WJ, Kozak R. More attention must be paid: the neurobiology of attentional effort. Brain research reviews. 2006;51(2):145-60.
73. Mather M, Harley CW. The locus coeruleus: Essential for maintaining cognitive function and the aging brain. Trends in Cognitive Sciences. 2016;20(3):214-26.
74. Alvarez GA, Cavanagh PJPs. The capacity of visual short-term memory is set both by visual information load and by number of objects. 2004;15(2):106-11.
75. Franconeri SL, Alvarez GA, Enns JT. How many locations can be selected at once? Journal of Experimental Psychology: Human Perception and Performance. 2007;33(5):1003.
76. Lapierre MD, Cropper SJ, Howe PD. Shared processing in multiple object tracking and visual working memory in the absence of response order and task order confounds. PloS one. 2017;12(4):e0175736.
77. Fougnie D, Marois R. Distinct capacity limits for attention and working memory: Evidence from attentive tracking and visual working memory paradigms. Psychological Science. 2006;17(6):526-34.
78. Marois R, Ivanoff J. Capacity limits of information processing in the brain. Trends in cognitive sciences. 2005;9(6):296-305.

79. Chesney DL, Haladjian HH. Evidence for a shared mechanism used in multiple-object tracking and subitizing. Attention, Perception, & Psychophysics. 2011;73(8):2457-80.
80. Xu Y, Chun MM. Selecting and perceiving multiple visual objects. Trends in cognitive sciences. 2009;13(4):167-74.

	Tracking	**Load**	**Crowding**
Frontal			
Supplementary motor area	+		+
Precentral gyrus	+		+
Middle frontal gyrus (including Frontal eye fields)	+		+ (L)
Middle frontal gyrus	+	+ (R)	+ (R)
Inferior frontal gyrus		+ (L)	
Medial frontal gyrus			+
Frontal medial orbital		-	-
Anterior Insula			+
Posterior insula	-		
Parietal			
Superior parietal lobule	+		
Inferior parietal lobule	+	+	+
Middle/posterior cingulate cortex	-	-	
Temporal/occipitial			
Fusiform		+	+
Middle/superior temporal gyrus	-	-	
Precuneus	+		
Inferior occpitial gyrus	+		
Middle occpitial gyrus	+		+
Lingual gyrus			+
Calcarine	-	-	
Subcortical			
Thalamus (pulvinar)	+		
Midbrain	+		
Hippocampus		-	-

Table 1. Summary of brain areas presenting hemodynamic changes due to the different conditions: Tracking > passive viewing (**Tracking**), Load 4 > Load 2 (**Load**), and High > Low crowding (**Crowding**). Positive activations are marked with a plus sign (+), and negative with a minus sign (-). If activations were lateralized, they are indicated as L = left, R = right. For precise coordinates, please refer to Tables 2 and 3.