Title

Working memory capacity of crows and monkeys arises from similar neuronal computations

Lukas Alexander Hahn[1], Dmitry Balakhonov[1], Erica Fongaro[1], Andreas Nieder[2] & Jonas Rose[1]

[1] *Neural Basis of Learning, Institute of Cognitive Neuroscience, Faculty of Psychology, Ruhr University Bochum, 44801 Bochum, Germany*

[2] *Animal Physiology, Institute of Neurobiology, University of Tübingen, 72076 Tübingen, Germany*

Keywords

Divisive normalization; computation principle; working memory capacity; capacity limitation; single-cell recordings; NCL; PFC; comparative cognition; flexible model of WM

Abstract

Complex cognition relies on flexible working memory, which is severely limited in its capacity. The neuronal computations underlying these capacity limits have been extensively studied in humans and in monkeys, resulting in competing theoretical models. We probed the working memory capacity of crows (*Corvus corone*) in a change detection task, developed for monkeys (*Macaca mulatta*), while we performed extracellular recordings of the prefrontal-like area nidopallium caudolaterale. We found that neuronal encoding and maintenance of information were affected by item load, in a way that is virtually identical to results obtained from monkey prefrontal cortex. Contemporary neurophysiological models of working memory employ divisive normalization as an important mechanism that may result in the capacity limitation. As these models are usually conceptualized and tested in an exclusively mammalian context, it remains unclear if they fully capture a general concept of working memory or if they are restricted to the mammalian neocortex. Here we report that carrion crows and macaque monkeys share divisive normalization as a neuronal computation that is in line with mammalian models. This indicates that computational models of working memory developed in the mammalian cortex can also apply to non-cortical associative brain regions of birds.

Introduction

Working memory (WM) can hold information for a short period of time to allow further processing in the absence of sensory input (*Cowan, 2017; Oberauer et al., 2018*). By bridging this gap between the immediate sensory environment and behavior, WM is a keystone for complex cognition. It is a very flexible memory system, yet severely limited in its capacity. While this capacity is often seen as a general cognitive bottleneck, for simple stimuli, like colors, the capacity is very similar between humans, monkeys, and crows (*Balakhonov and Rose, 2017; Buschman et al., 2011; Cowan, 2001; Luck and Vogel, 1997*).

Different models have been proposed to conceptualize how this capacity limit arises. This work motivated many psychophysical and electrophysiological experiments that in turn led to a spectrum of more refined models of WM (*Ma et al., 2014*). 'Discrete models' of WM argue that a fixed number of items can be stored. Once this capacity is reached, an additional item can only be maintained if it replaces a previous item (*Awh et al., 2007; Fukuda et al., 2010; Luck and Vogel, 1997; Vogel and Machizawa, 2004; Zhang and Luck, 2008*). 'Continuous models' describe WM as a flexible resource that is allocated to individual items. A minimum amount of this resource has to be allocated to each item for successful retention, thereby resulting in a capacity limit *(Bays and Husain, 2008; Berg et al., 2012; Wilken and Ma, 2004*).

On the neurophysiological level, models of WM capacity suggest that interference between memory representations ('*items*') within the neuronal network is a source of information loss and capacity limitation (*Bouchacourt and Buschman, 2019; Lundqvist et al., 2016, 2011; Schneegans et al., 2020*). Interference may arise due to divisive normalization that appears as competition between items, related to oscillatory dynamics (*Lundqvist et al., 2016, 2011*), WM flexibility (*Bouchacourt and Buschman, 2019*), and neuronal information sampling (*Schneegans et al., 2020*). Divisive normalization is a computational principle that acts upon neurons when presenting multiple stimuli simultaneously, it normalizes neuronal responses by creating 'a ratio between the response of an individual neuron and the summed activity of a pool of neurons' (*Carandini and Heeger, 2012, p. 51*). An effect related to divisive normalization can be observed when two stimuli are presented either individually or simultaneously within the receptive field of a visual sensory neuron. The neuron's firing rate when the stimuli are presented simultaneously becomes normalized by the populations' responses to each individual stimulus (*Carandini et al., 1997; Heeger, 1992*). This effect also occurs in relation to attentive processes (*Reynolds et al., 1999; Reynolds and Heeger, 2009*). Normalization of neuronal responses is commonly observed in many species throughout the animal kingdom, not just in sensory, but also in cognitive domains

(*Carandini and Heeger, 2012*).

Investigations into WM capacity and model predictions focus mostly on humans and monkeys. By extending this work to include birds, one can gain a unique comparative perspective. Crows have a similar limit in WM capacity and neuronal correlates of WM are comparable to monkeys' (*Balakhonov and Rose, 2017; Nieder, 2017*). But while the neuronal architecture of sensory areas is similar between birds and mammals, higher associative areas, critical for WM, do not share a common architecture between the species (*Stacho et al., 2020*). Therefore, an outstanding question is whether modern models of WM such as the 'flexible model' capture WM capacity in general, or if their predictions (e.g., divisive normalization) are confined to the mammalian neocortex. To resolve this, it is crucial to investigate the avian brain to understand how its different organization can produce such similar behavioral and neurophysiological results. While the neuronal correlates of WM maintenance in birds have been investigated in some detail (*Diekamp et al., 2002; Hartmann et al., 2018; Rinnert et al., 2019; Rose and Colombo, 2005; Veit et al., 2014*), a neurophysiological investigation of WM capacity limitation is still lacking. The avian forebrain structure, *nidopallium caudolaterale* (NCL) is a critical component of avian WM. The NCL is considered functionally equivalent to the mammalian prefrontal cortex (PFC) *(Güntürkün and Bugnyar, 2016; Nieder, 2017*), as it receives projections from all sensory modalities (*Kröner and Güntürkün, 1999*), projects to premotor areas (*Kröner and Güntürkün, 1999*), and is a target of dopaminergic innervation (*Waldmann and Güntürkün, 1993*).

To investigate the neurophysiology of WM capacity in birds, we adopted a task design developed for monkeys (*Buschman et al., 2011*) to use it with carrion crows (*Corvus corone*). Our animals were trained to memorize an array of colors and to indicate which color had changed after a short memory delay, while we performed extracellular recordings of individual neurons in the NCL using multichannel probes. We expected to find a clear correlate of WM representations in NCL neurons and a load-dependent response modulation based on divisive normalization of neuronal responses. This would allow us to evaluate if the behavioral WM capacity observations of crows fit a 'discrete' or 'continuous' WM resource model. If the neuronal responses also fit the contemporary neurophysiological models of WM capacity limitations (*Bouchacourt and Buschman, 2019; Lundqvist et al., 2016, 2011; Schneegans et al., 2020*) it would further suggest that crows and monkeys have convergently evolved a similar neurophysiological basis for WM capacity despite a different architecture of the critical forebrain structures.

Materials and Methods

Subjects.

Two hand-raised carrion crows (*Corvus corone*) of 2 years of age served as subjects in this study. The birds were housed in spacious aviaries in social groups. During the experimental procedures, the animals were held on a controlled food protocol with *ad libitum* access to water and grit. All experimental procedures and housing conditions were carried out in accordance with the National Institutes of Health *Guide for Care and Use of Laboratory Animals* and were authorized by the national authority (*Regierungspräsidium*).

Experimental setup.

We used operant training chambers (50 x 50.5 x 77.5 cm, width x depth x height) equipped with an acoustic pulse touchscreen (22'', ELO 2200 L APR, Elo Touch Solutions Inc, USA) and an infrared camera (Sygonix, Nürnberg, Germany) for remote monitoring. The birds sat on a wooden perch so that the distance between the bird's eye and the touchscreen was 8 cm. Food pellets were delivered as a reward via a custom-made automatic feeder (plans available at www.jonasrose.net). The position of the animal's head was tracked online during the experiment by two open-source computer vision cameras ('Pixy', CMUcam5, Charmed Labs, Texas, USA) that reported the location and angle between two LEDs. For tracking, we surgically implanted a lightweight head-post and used a lightweight 3D-printed mount with LEDs that was removed after each experimental session. The system reported the head-location at a frame rate of 50 Hz and data was smoothed by integrating over 2 frames in Matlab using custom programs on a control PC. All experiments were controlled by custom programs in Matlab using the Biopsychology (*Rose et al., 2008*) and Psychophysics toolboxes (*Brainard, 1997*). Digital input and output of the control PC were handled by a microcontroller (ODROID C1, Hardkernel co. Ltd, Anyang, South Korea) connected through a gigabit network running custom software (available at: www.jonasrose.net).

Behavioral protocol.

The behavioral protocol was identical to the one described in (Balakhonov and Rose, 2017). We trained the birds to perform a delayed change localization task that had previously been used to test the performance under different working memory loads in primates (*Buschman et al., 2011*). Each trial started after a 2 s inter-trial-interval, with the presentation of a red dot centered on the touchscreen (for a maximum of 40 s). The animals initiated the trial by centering their head in

front of the red dot for 160 ms. This caused the red dot to disappear and a stimulus array of two to five colored squares to appear (Fig. 1A). The colored squares were presented for a period of 800 ms, during which the animals had to hold their head still and center their gaze on the screen (no more than 2 cm horizontal or vertical displacement, and no more than 20° horizontal or vertical rotation). Failure to hold the head in this position resulted in an aborted trial. This sample phase was followed by a memory delay of 1000 ms after which the stimulus array reappeared with one color exchanged. The animal had to indicate the location of the color change by pecking the respective square. Correct responses were rewarded probabilistically (BEO special pellets, in 55 % of correct trials, additional 2 s illumination of the food receptacle in 100 % of correct trials). Incorrect responses to colors that had not changed or a failure to respond within 4 s resulted in a brief screen flash and a 10 s timeout.

The stimuli were presented at six fixed locations on the screen (1 – 6, Fig. 1A). For each location, a unique color pair was randomly chosen from a set of 14 colors (two possible pairs of colors were excluded due to similarity). Thus, during any given experimental session, a random pair of colors was fixed to each of the six locations. The order of presentation of colors within a pair, the target location (where the color change occurred), and the number of stimuli in the array (two to five) were randomized and balanced across trials so that each condition had an equal likelihood to appear. The color squares had a width of 10 degrees of visual angle (DVA) and were placed on the horizontal meridian of the screen and at 45.8 DVA above or below the meridian at a distance of 54 and 55.4 DVA from the center. This arrangement in combination with the head tracking ensured that all stimuli appeared outside of the binocular visual field of crows (37.6 DVA (*Troscianko et al., 2012*)).

Surgery.

Both animals were chronically implanted with a lightweight head-post to attach a small LED-holder during the experiments. Before surgery, animals were deeply anesthetized with ketamine (50 mg/kg) and xylazine (5 mg/kg). Once deeply anesthetized, animals were placed in a stereotaxic frame. After attaching the small head-post with dental acrylic, a microdrive with a multi-channel microelectrode was stereotactically implanted at the craniotomy (Neuronexus Technologies Inc., Ann Arbor MI, DDrive). The electrode was positioned in NCL (AP 5.0, ML 13.0) of the left hemisphere (coordinates for the region based on histological studies on the localization of NCL in crows (*Veit and Nieder, 2013*)). After the surgery, the crows received analgesics.

Electrophysiological recordings.

Extracellular single neuron recordings were performed using chronically implanted multi-channel microelectrodes. The distance between recording sites was 50 µm. The signal was amplified, filtered, and digitized using Intan RHD2000 headstages and a USB-Interface board (Intan Technologies LLC, Los Angeles CA). The system also recorded digital event-codes that were sent from the behavioral control PC using a custom IO-device (details available at www.jonasrose.net). Before each recording session, the electrodes were advanced manually using the microdrive. Recordings were started 20 minutes after the advancement, and each recording site was manually checked for neuronal signals. The signals were recorded at a sampling rate of 30 kHz and filtered with a band-pass filter at recording (0.5 kHz – 7.5 kHz). The recorded neuronal signals were not pre-selected for task involvement.

We performed spike sorting using the semi-automatic Klusta-suite software (*Rossant et al., 2016*), which uses the high electrode count and their close spacing to isolate signals of single neurons. For spike sorting, we filtered with a low-pass of 500 Hz and a high pass of 7125 Hz. The software utilizes the spatial distribution of the recorded signal along the different recording sites to untangle overlapping signals and separate signals with similar waveforms but different recording depths.

Data analysis.

All statistical analyses were performed in Matlab (2018b, Mathworks Inc.) using commercially available toolboxes (Curve Fitting Toolbox Version 3.5.3, Statistics and Machine Learning Toolbox Version 10.2) and custom code. For all statistical tests, we assumed a significance level of $\alpha = 0.05$, unless stated otherwise. Trials were classified as error trials if the bird chose a location where no change of colors had appeared. Trials in which the bird did not choose any location or failed to maintain head fixation were not analyzed. All correct trials were included in the analysis of neural data. Because there were only very few error trials in the load 1 condition, we performed error trial analysis only for the load 2 and load 3 conditions.

The behavioral data were analyzed as described in our previous study (*Balakhonov and Rose, 2017*), estimating the working memory capacity K for each load by equation 1.

Equation 1: $$K = n * p$$

Where p is the percentage correct and n is the number of items in working memory. This estimate has been applied to similar primate data and in studies with humans (*Johnson et al., 2013; Kornblith et al., 2016*).

Information about color identity.

Based on a one-way ANOVA of color identity at a given location, we calculated a percent explained variance statistic (PEV) to measure the effect size of neuronal modulation. Its main parameter ω^2 is a measurement for the percentage to which the tested factor can explain the variance of the data, and it is calculated from the sum of squares of the effect (SS_{effect}) and the mean squares of the within-group (error) variance (MS_{error}) (Eq. 2).

Equation 2: $$\omega^2 = \frac{SS_{effect} - df * MS_{error}}{SS_{total} + MS_{error}}$$

For each neuron, we determined a 'favorite location', which was defined as the location with the highest cumulative PEV, contralateral to the electrode position, across four non-overlapping bins during the sample phase (bin size 200 ms, advanced in steps of 200 ms, from start till the end of the sample phase). The significance of calculated effect size values was determined by a permutation test. We ran the permutation to calculate the likelihood of getting an explained variance value bigger than the one calculated from the actual distribution of the data by randomly permuting the color identity labels and calculating the PEV 1000 times. The test thereby does not assume any distribution of the data and returns an unbiased estimate of the likelihood of generating an effect size within the data randomly. The measured value of explained variance from the actual dataset was assumed to be significant if the likelihood of randomly generating a bigger value was below 5 %.

We tested the proportions of significant neurons we found for the different trial phases by performing a binomial test, assuming a significance level α = 0.05 (Eq. 3).

Equation 3: $$P(X = i) = B(p_0, n) = (n\,k)p_0^i(1 - p_0)^{n-i}$$

Calculating the probability *P*, of finding *X* significant neurons, given a total amount of *i* (362) neurons, and a probability p_0 of 5 % finding a significance by chance.

Population analyses.

We considered neuronal significance (i.e., significant PEV as determined above) for each load independently. This means, we tested if the PEV of a neuron was significant three times with the permutation method described above: once for each of the three load conditions. Therefore, we can report seven groups of significance (Tab. 1, Fig. 3C). Subsequently, we created three pooled groups (Tab. 1) from all neurons with a significant PEV at each individual load. We used these pooled groups for the population analyses (Figs. 4 & 5). Neurons of these pooled groups, with a significant PEV during the sample phase were assigned to the 'sample-population', and neurons

with a significant amount of information during the memory-delay phase were assigned to the 'delay-population' (significance criterion: one significant 200 ms bin, at $\alpha = 0.005$). Thus, neurons with significant PEV during both the sample and delay phase were included in both subpopulations.

We corrected for the unequal amount of correct and error trials when comparing information about color (PEV) between the trial conditions, by sub-sampling correct trials with the number of error trials 1000 times for each neuron. The resulting PEV-values of correct trials were then averaged for each neuron, this population of averaged PEV values was then statistically tested against the PEV values of error trials (of the same neurons) using a dependent t-test.

Table 1: Overview of significant groups. The '+' denotes that a neuron of the respective group had a significant PEV in the respective load condition. The '-' denotes that a neuron of the respective group did not have a significant PEV in the load respective condition. The pooled groups contained only neurons with a '+' for the respective load condition.

Load 1	Load 2	Load 3	Group name	
+	**-**	**-**	Load 1 neurons	Group I
-	**+**	**-**	Load 2 neurons	Group II
-	**-**	**+**	Load 3 neurons	Group III
+	**+**	**-**	Load 1&2 neurons	Group IV
+	**-**	**+**	Load 1&3 neurons	Group V
-	**+**	**+**	Load 2&3 neurons	Group VI
+	**+**	**+**	Load 1&2&3 neurons	Group VII
Pooled group 1	**Pooled group 2**	**Pooled group 3**		

Divisive normalization like regularization.

We tested for the presence of divisive normalization using the method of (Reynolds et al., 1999). Three conditions were considered: (1) neuronal response to stimulus A, (2) neuronal response to stimulus B, and (3) neuronal response to the simultaneity of stimulus A and B. As we wanted to relate this to the information about color identity, we selected subsets of the favorite location and the additional two ipsilateral locations. To test how the neurons altered their response when multiple stimuli were presented simultaneously, we calculated the color selectivity index (SE) and the sensory interaction index (SI) of each neuron.

SE_i was calculated by subtracting the normalized firing rate for the chosen reference color *i* (REF_i)

at the neuron's favorite location, from a second color *j* ($PROBE_j$) at a different location (ipsilateral to the favorite location, Eq. 4).

Equation 4: $$SE_i = PROBE_j - REF_i$$

The resulting selectivity index lies between -1 (completely selective for the reference color) and 1 (completely selective for the probe color). SI was calculated (Eq. 5) by subtracting the normalized firing rate for REF_i from the normalized firing rate of the combination of REF_i and $PROBE_j$ ($PAIR_{i,j}$).

Equation 5: $$SI_{i,j} = PAIR_{i,j} - REF_i$$

This interaction index also lies between -1 (full suppression of reference stimulus by the probe stimulus) and 1 (full enhancement of the reference stimulus by the probe stimulus). As each of the three locations had two possible colors, we calculated eight SE and SI indices per neuron and performed a linear regression for all indices. This is required as each stimulus combination is informative about the normalization.

The effects of divisive normalization were compared between the sample and the delay phase. Therefore, SE and SI indices were calculated across the entire sample (800 ms) and memory delay (1000 ms) phase. Neurons with significant information were accordingly identified over the entire sample and delay as one bin, using the permutation test described in the section '*information about color identity*'.

Hierarchical clustering.

To visualize the different groups of neurons that encoded and maintained information about the color identity during different phases of the trial we performed a hierarchical clustering analysis in Matlab on the normalized PEV values of individual neurons throughout the trial. We used a (1 - correlation) distance metric and an average distance linkage function for a maximum of seven clusters. The maximum number of clusters was first determined by calculating the clustering for different amounts of clusters (1 to 10) and subsequently calculating the within-cluster sum-of-squares. This resulted in a graph that allowed us to visually inspect the tradeoff between cluster number and fit-improvement, from which we estimated the inflection point (elbow-method). A cluster number of seven presented the best tradeoff that allowed visualization of the different groups at an acceptable clustering success. We then ordered the neuron clusters to minimize the average distance between the clusters in the dendrogram.

Results

The working memory capacity of crows is similar to that of monkeys.

The behavioral performance was influenced by the number of colored squares on the screen. It significantly decreased with an increasing number of ipsilateral squares (median performances, load 1: 95.88 %, load 2: 78.31 %, load 3: 58.21 %; Friedman test: $X^2 = 92.00$, $p < .001$, Fig. 1B). We ran a generalized linear model with ipsilateral load (i.e. load of hemi-field where a color changed), contralateral load (i.e. load of hemi-field without a color change) and their interaction as predictors for performance ($R^2_{adj} = .78$, $F(460,456) = 555.00$, $p < .001$). We found that the number of ipsilateral colors significantly reduced performance ($\beta_{ipsi} = -.177$, $t(459) = -18.00$, $p < .001$), whereas the number of contralateral colors did not ($\beta_{contra} = -.021$, $t(459) = -1.77$, $p = .0772$; Fig. 1C). There was also a significant interaction between ipsilateral and contralateral load ($\beta = -.024$, $t(458) = -4.28$, $p < .001$). We compared this model to a reduced model, where we omitted the non-significant β_{contra} and found that this reduced model ($R^2_{adj} = .78$, $F(460,457) = 828.00$, $p < .001$) explained the performance equally as well ($|\Delta LLR| = .0102$). Therefore, we conclude that contralateral load by itself did not significantly affect performance. We calculated the capacity K (see methods) for all full WM-loads (i.e., two to five items). The capacity K peaks at four items (mean +/- SEM: 3.05 +/- .038, Supplementary Fig.1). These observations are very similar to observations made in primates (*Buschman et al., 2011*) and fully reproduce our earlier behavioral findings (*Balakhonov and Rose, 2017*).

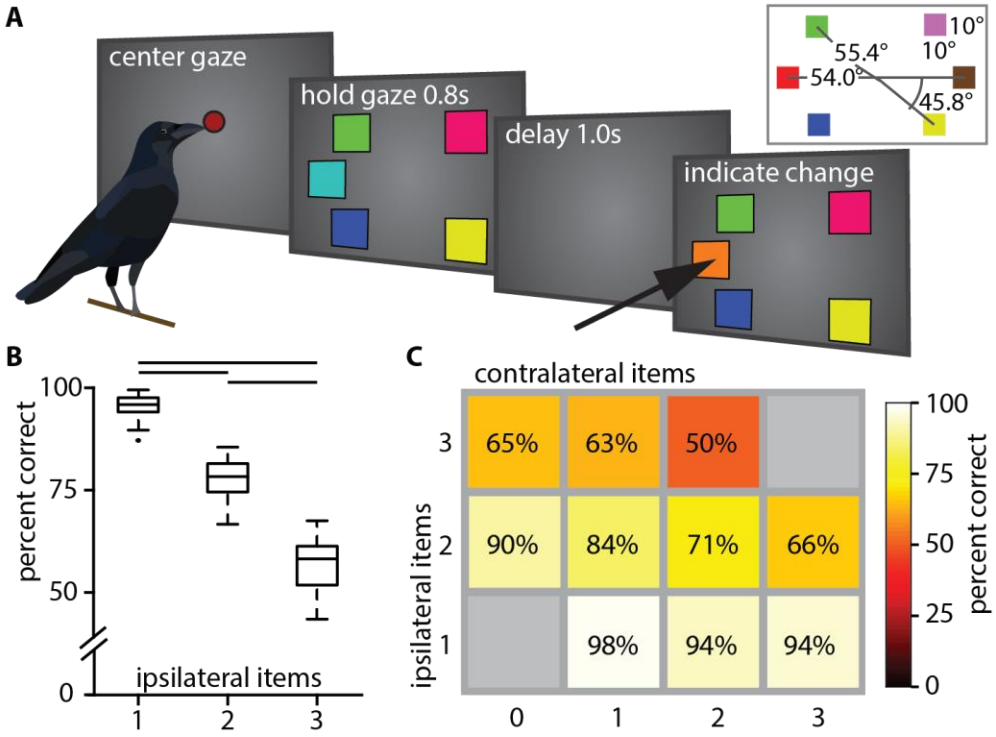


Figure 1: (A) Behavioral paradigm (reproduced from Balakhonov & Rose, 2017). (B) Boxplot of performance for different ipsilateral loads. Horizontal lines indicate significant differences between loads, box indicates the median, 1st, and 3rd quartile (whiskers extend to 1.5 times the inter-quartile range). (C) Mean performance matrix for ipsi- and contralateral load combinations. Additional contralateral items at an ipsilateral load of 1 barely affected performance (bottom row). At higher ipsilateral loads additional contralateral items reduced performance more clearly (middle and top row). Statistical modeling revealed an interaction at these higher loads (see text).

Neurons of the NCL encode the color identity and maintain it in working memory.

We recorded 362 neurons from the NCL of two crows performing the WM task (delayed change localization). All reported effects were also present in each individual bird (Supplementary Figs. 2 & 3), we, therefore, pooled the data for population analysis. A large subset of neurons responded to the presence of a color (i.e. at load 1) by substantially increasing or decreasing their firing rate relative to baseline. This change in firing rate occurred selectively, depending on the presented color either in the sample (Fig. 2A) or the delay period (Supplementary Fig. 4). For most neurons, this difference in firing rate between the two possible colors became attenuated when the load increased from one to two colors, and it was further attenuated from two to three colors. To quantify this effect, we calculated the amount of information about the color identity at a neurons' favorite location as the percent explained variance (PEV) during a memory load of one, two, or three items in bins of 200 ms (see methods for details). Most neurons did not sustain information about color (measured as a significant PEV, henceforth '*information*') throughout the

entire sample or memory delay but rather had shorter periods in which the information was significant (Fig. 2A bottom).

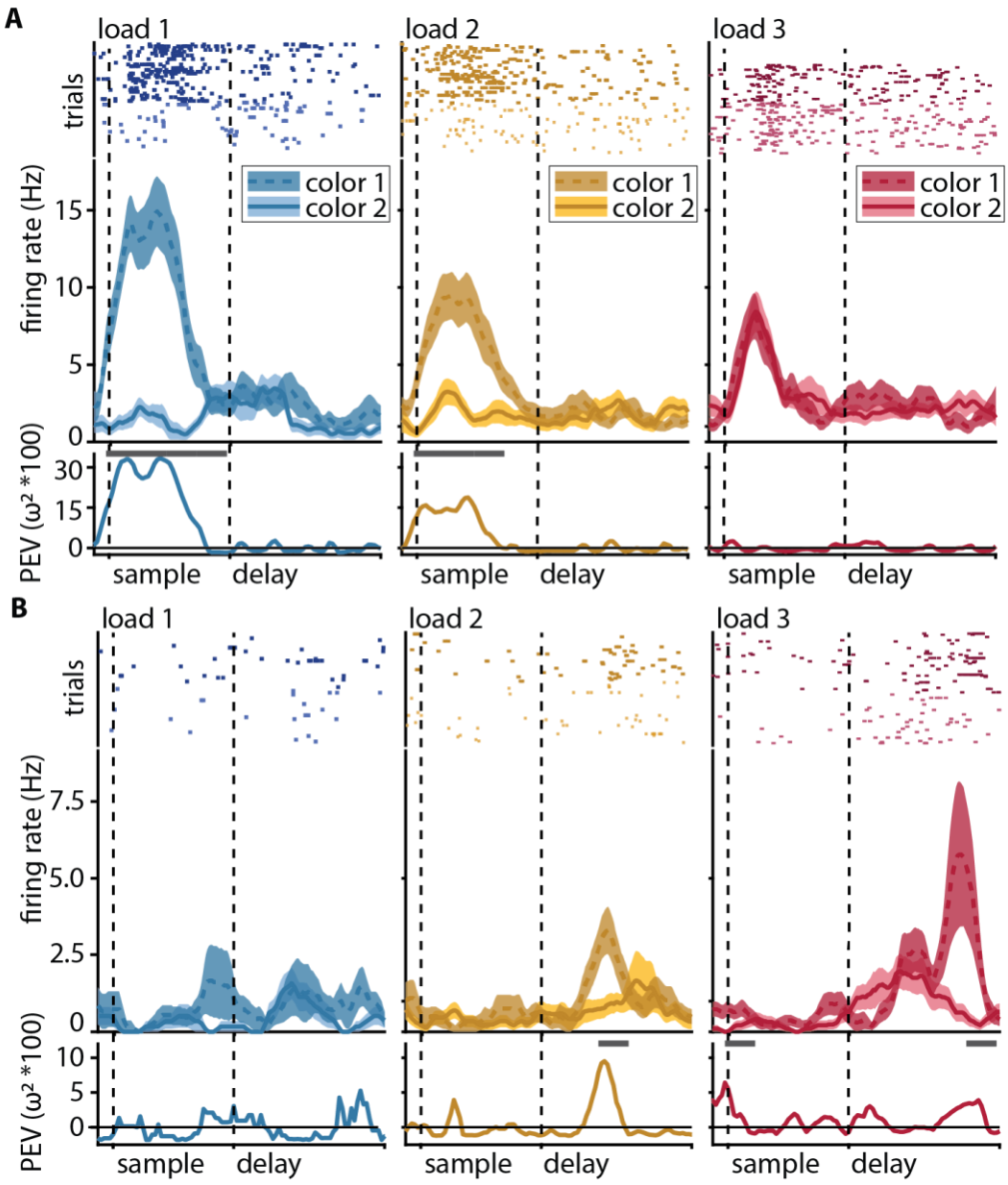


Figure 2: Color discrimination in the neuronal response (information, PEV) generally decreases with load, but some neurons show the opposite effect. (A) Example of a sample neuron with color information decline at load 1 (blue), load 2 (yellow), and load 3 (red). Top: raster plot, where every dot represents a single spike during the individual trials (rows of dots); middle: peri-stimulus-time histogram (PSTH) of average firing rate (solid line for color ID 1, dashed line for color ID 2) with the standard error of the mean (shaded areas); bottom: percent explained variance of color identity (a measure of information about color) along the trial, the line at the top of the y-axis indicates significant bins. (B) Same as in (A) for an example of a delay neuron with information gain at a higher load.

To better capture the time points when the individual neurons carried information, we performed a hierarchical clustering analysis of the PEV values of the individual neurons at load 1 (see methods for details). We found a total of seven clusters that were organized into two overarching groups (Fig. 3A). Group 1 contained neurons (n = 227) that showed peak information during the sample and early delay phase, while group 2 contained neurons (n = 135) that showed peak information during the delay phase. For each neuron, we then calculated if it carried a significant amount of color information by applying a permutation test (for all bins at load 1, see methods). The individual neurons were then further classified into three groups depending on the phase in which they had a significant amount of information (Fig. 3B). Overall, 37.57 % (n = 136) of neurons were significant during the sample phase, 9.39 % (n = 34) of neurons were significant during the memory delay, and 14.64 % (n = 53) of neurons were significant during both the sample phase and the memory delay (all proportions of neurons were significantly higher than expected by chance (binomial test, see methods, all $p < .001$)). Refer to Fig. 2A for an example neuron, significant at load 1 with a large differentiation in firing rate between color identities (a large PEV) and a loss of differentiation with increasing ipsilateral load.

Further inspection of individual neuronal activity revealed, however, that a substantial number of neurons responded differently. Instead of losing information at higher loads, many neurons gained information (e.g., Fig. 2B, Supplementary Fig. 5). Thus, we additionally performed the permutation testing for loads 2 and 3 to determine which neurons had significant information (see methods). We found that many of the neurons that did not have significant information at load 1 did have significant information at load 2 and load 3 (Fig. 3C). For the memory delay, more than half of the significant neurons we detected were only significant for either load 2 or load 3, compared to only 36 % of neurons that were significant at load 1 (Fig. 3C middle). By including the higher loads in our analysis, we found a total of 249 (68.78 %) sample neurons and 94 (25.97 %) delay neurons. For the population analyses, we subsequently pooled all significant neurons into three groups (one per load). These pooled groups were then each subdivided into sample and delay neurons (i.e., ‘sample-load1’, ‘delay-load1’, ‘sample-load2’, etc., see table 1 in the methods for an overview).

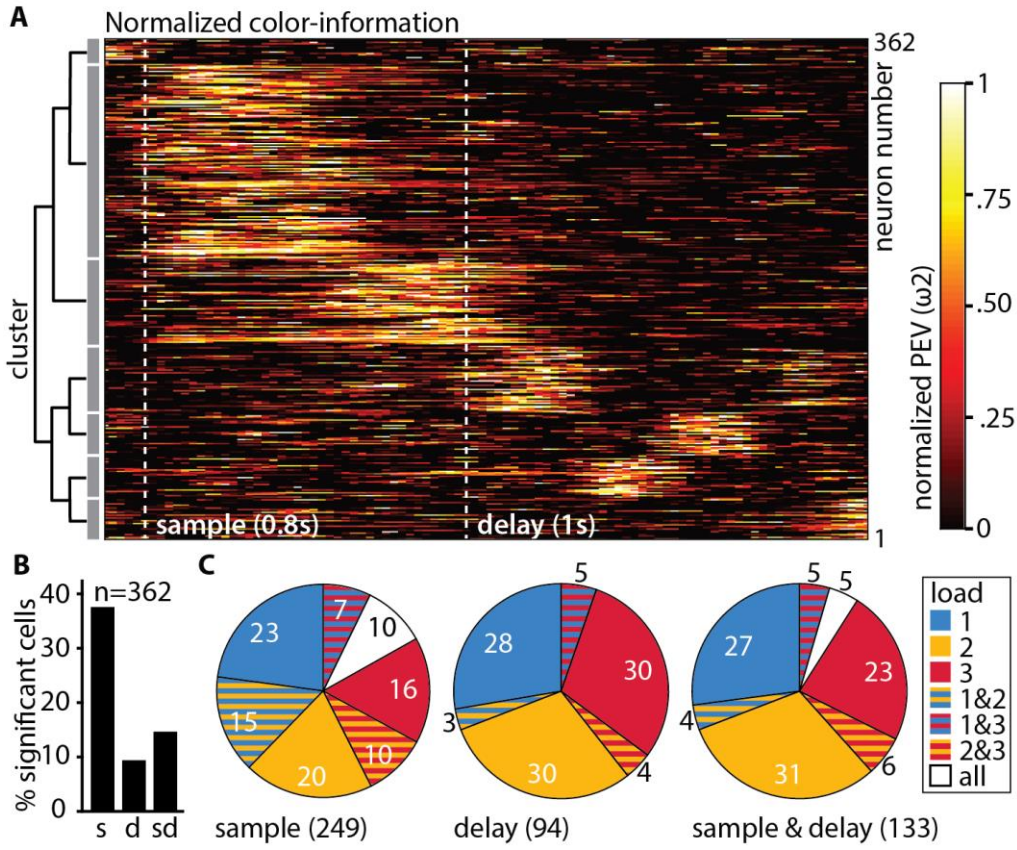


Figure 3: (A) The neuronal population can be best described by 7 individual clusters. (B) Percentages of neurons (total n = 362) with significant color information at load 1, during the sample and the delay. (C) Percentages (rounded) of significant neurons in individual load conditions for sample (n = 249), delay (n = 94), and sample & delay (n = 133). The pieces of the pies depicting significance at a specific load relate to the number of significant neurons in the respective phase (e.g. 36 % of the 94 delay neurons (i.e., 34 neurons) are significant at load 1 (all pieces contain blue), these are the same neurons that make up the 9.36 % of the total 362 neurons depicted in B).

The neuronal population has gradually less information with increasing load.

The clustering analysis indicated that the population of neurons as a whole did sustain the color information throughout the entire trial (Fig. 3A). Plotting the information averages of each of the three 'sample-populations' and the 'delay-populations' over time confirmed this result (Fig. 4A, Supplementary Fig. 6). After the onset of the stimulus array, the average information exhibited a sharp increase that peaked roughly 400 ms after stimulus onset and remained at an elevated level throughout the memory delay, until the choice array appeared. Results obtained from neurons of the lateral prefrontal cortex of monkeys indicated distinct hemispheric independence of WM capacity (*Buschman et al., 2011*). This means that increasing ipsilateral load (i.e. load in the hemifield containing the target for which information is assessed) should affect neuronal

processing while increasing contralateral load should not. This effect might be further emphasized in birds due to the full decussation of their optic nerve (*Husband and Shimizu, 2001*). Parallel to the behavioral results and in line with the results from monkeys, we found a strong effect of ipsilateral load on the information maintenance, as there was a sharp drop in information when the load increased from 1 item to 2 items (Fig. 4A, blue and yellow curves). The addition of a third item only slightly decreased the maintained information further (Fig. 4A, red curve). The load dependence was much more pronounced during the sample period than during the memory delay where the information remained at a lower elevated level. Notably, the load effect was only present for ipsilateral manipulations. If the number of items on the contralateral side was increased, the information encoded about the colors at the favorite location did not change (Fig. 4A, right). To compare our results to the results obtained in monkeys we also applied the method of (Buschman et al., 2011) for testing the ipsilateral load effect during the sample and delay phase, by splitting each phase into an early and a late portion (first and second 400 ms of the sample, and first and second 500 ms for the delay). We did find a significant drop in information with an increase in load from 1 through 3 in the early ($F(2,537) = 18.73$, $p < .001$, $\omega^2 = 0.0616$) and late ($F(2,536) = 20.07$, $p < .001$, $\omega^2 = 0.0661$) sample period and the early ($F(2,267) = 6.88$, $p = .0012$, , $\omega^2 = 0.0417$) and late ($F(2,267) = 3.85$, $p = .0225$, , $\omega^2 = 0.0207$) delay period (Fig. 4B). There was a large and significant drop between 1 and 2 items (post hoc Bonferroni corrected multiple comparisons: early and late sample $p < .001$, early delay $p < .001$, late delay p = 0.0198) and 1 and 3 items (post hoc Bonferroni corrected multiple comparisons: early and late sample $p < .001$, early delay $p = 0.019$, late delay p > 0.05) but no difference between loads 2 and 3 (post hoc Bonferroni corrected multiple comparisons: all $p > 0.05$). The maintenance of a significant amount of information at higher loads (even for 3 items, early sample $t(156) = 7.55$, $p < .001$; late sample $t(156) = 8.73$, $p < .001$; early delay $t(87) = 3.84$, $p < .001$; late delay $t(87) = 4.73$, $p < .001$) and its gradual reduction when items were added to the corresponding hemifield are indicative of a flexible resource allocation and not an all-or-nothing slot-like WM. Furthermore, if there is a flexible resource, in error trials a small but insufficient amount of resource might still be allocated to an item.

Indeed, error trial analysis (applying correct trial sub-sampling, see methods) for the load 2 & 3 conditions further supported this interpretation. The amount of information in the early and late sample phase remained above zero (load 2: early, $t(186) = 3.25$, $p = 0.0014$; late, $t(186) = 5.33$, $p < .001$; load 3: $t(156) = 4.21$, $p < .001$; Fig. 4B asterisks), and was significantly smaller than in correct trials (load 2: late, $t(186) = 2.81$, $p = .0055$, $d = 0.26$; load 3: late $t(156) = 2.55$, $p = .0117$, $d = 0.23$). Additionally, there was no further maintenance throughout the memory delay at any

load (Fig. 4B, PEV at load 2 and 3 in error trials delay, all non-significant). This indicates that a failure to report which color had changed at higher loads (2 and 3 ipsilateral items) resulted from a smaller amount of information encoding during the sample phase that was not maintained through the delay.

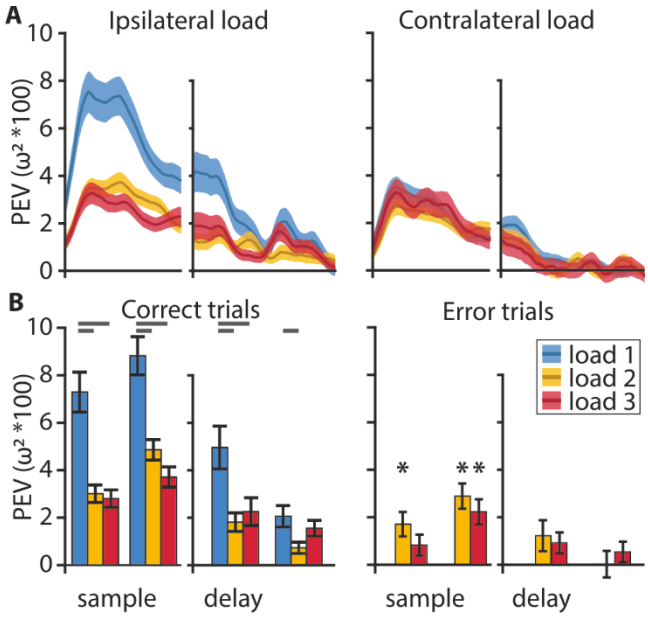


Figure 4: Information encoding at the population level. (A) Color information (PEV) decreases with an increasing ipsilateral load but not with an increasing contralateral load. (B) On correct trials (left) color is represented during the early and late phase of the sample and, to a lesser degree, during the early and late delay. On error trials (right), color information can be found in the early sample phase at load 2, and in the late sample phase at loads 2 & 3 (asterisks). Analysis of load 1 error trials was omitted due to their very low abundance. Statistical comparisons of correct vs. error trial information were performed on sub-sampled correct trials. Early and late sample each 400 ms, early and late delay each 500 ms, shaded areas, and error bars indicate the standard error of the mean.

Higher loads produce divisive normalization-like neuronal responses.

We next wanted to understand the neuronal mechanisms behind the information loss at higher WM loads. For that, we analyzed how the responses of individual sample- and delay-neurons changed when the load increased from 1 color to 2 colors. For the 'sample-populations' and the 'delay-populations', an increasing number of items reduced the amount of encoded information about the color identity (Fig. 4). This effect was due to neurons that had a large difference of firing rates between the color 1 and color 2 at load 1 (high PEV, i.e. information about color), and reduced differentiation at load 2 (small PEV, no or little information about color, e.g. Fig. 2A). 'Divisive-normalization-like regularization' (DNR, (*Carandini and Heeger, 2012*)) can explain this effect. DNR describes the computation that takes place when two stimuli are presented

simultaneously. In a simplified case a neuronal response becomes normalized, analogous to vector normalization, with a normalization factor consisting of the simultaneous stimuli (*Carandini and Heeger, 2012*). Applied to our context, a consequence of DNR would be a reduced differentiation between two color identities at load 2 because differences in firing rate (for each stimulus by itself) at load 1 would be normalized at load 2 (resulting in information loss, e.g., Fig. 2A). We, therefore, hypothesized that DNR was observable for neurons with significant information at load 1.

We tested for DNR in the NCL by calculating a selectivity index (SE) and a sensory interaction index (SI) for each neuron for the sample phase and the memory delay phase (*Reynolds et al., 1999*), see methods for details). SE indicates how strongly the neuronal response is driven by a color at the favorite location of the neuron (reference) in relation to a selected probe color (when either is presented alone). SI indicates how the probe color interacts with the reference color when both are presented simultaneously. Values of both indices, SE and SI, lie between -1 and +1. The addition of a probe color influences the response to the reference color by either suppressing the firing rate of the reference color (if the reference elicits a higher firing rate than the probe, i.e., SE < 0), or increasing the firing rate for the reference color (if the probe elicits a higher firing rate than the reference, i.e., SE > 0). If DNR was present, this influence to suppress or enhance neuronal responses should be an even mixture at the population level, resulting in a significant regression between SE and SI with a slope of around 0.5 (*Bouchacourt and Buschman, 2019*).

We compared regressions for the sample and delay phase (each as one bin, see methods for details) for two groups of neurons: information-carrying neurons (significant information at load 1), and non-informative neurons (no information at load 1 or at load 2; Supplementary Fig. 7). We found that DNR was present in both sample and delay phases (Fig. 5A). Information-carrying sample neurons had a fitted slope of 0.47 (R^2_{adj} = .39, $F(1,838)$ = 547.69, p < .001, CI = [0.43 0.51]) and delay neurons had a slope of 0.50 (R^2_{adj} = .34, $F(1,342)$ = 175.60, p < .001, CI = [0.43 0.58]). As the slopes were not significantly different from 0.5, this indicates that reference and probe color had an equal influence on neuronal responses. We thus show that DNR was observable in the neuronal population, and as a consequence of this computation, neurons had generally less information about the color identity at load 2.

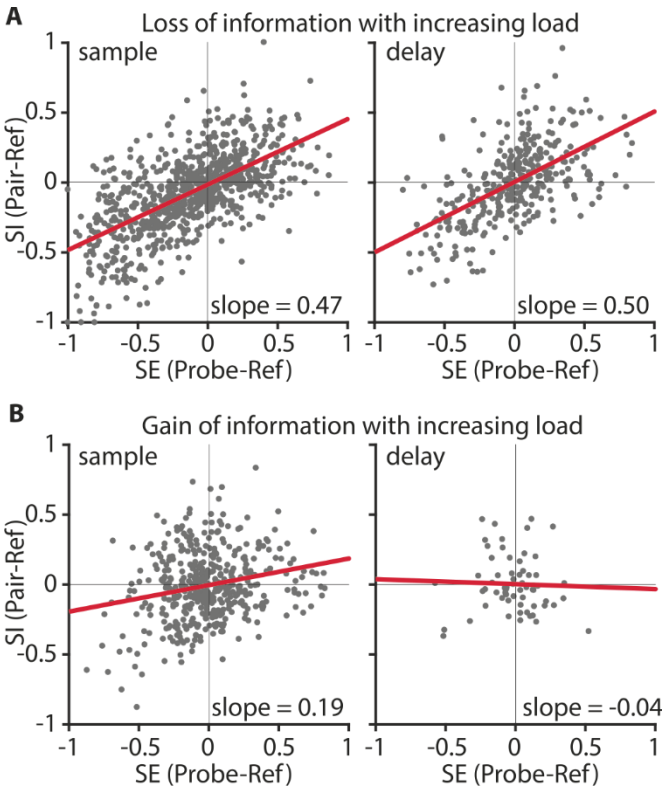


Figure 5: Divisive normalization-like regularization was observable for neuronal responses of neurons losing information (A) but not for neurons gaining color information at load 2 (B). Selectivity (SE) indicates how much the neuronal response is influenced by a color, relative to a second color when either is presented alone. Sensory interaction (SI) indicates how much the neuronal response is influenced by either color when both were displayed simultaneously. Slopes close to 0.5 indicate an equal influence of both colors. Slopes < 0.5, or > 0.5 indicate a weighted influence of a color. (A) Information carrying neurons in the sample (n = 105; left) and delay (n = 43; right) population. (B) Information gaining neurons in the sample (n = 56 ; left) and delay (n = 8; right) population. The red line indicates the regression fit.

Gain of information at load 2 can be explained by neuronal normalization.

Some neurons showed encoding of color identity at higher loads, instead of loss of information. These neurons were abundant in both the sample phase and the delay phase (Fig. 3C). For example, the neuron shown in figure 2B did not differentiate between color identities at load 1 but did so for load 2, thus, representing a case of information gain (instead of loss) at a higher load. We wanted to understand if DNR, the mechanism that we found reduced color information at load 2, could also produce color differentiation. This would be the case if the interaction between the additional color and the target color is unequal, because neurons without a color differentiation at load 1 may have gained differentiation at load 2 (e.g. if the interaction of probe color 1 with reference color 1 is larger than the interaction of probe color 1 with reference color 2). This would

result in a population regression slope smaller than 0.5. We thus hypothesized that the population of neurons showing information at load 2, but not at load 1 (e.g. Fig. 2B), would have a smaller slope than the neurons that lost information (Fig. 5A). Sample neurons had a slope of 0.19 (R^2_{adj} = .05, $F(1,446) = 23.0$, $p < .001$, CI = [0.11 0.27], Fig. 5B), and delay neurons had a slope of -0.04 (R^2_{adj} = -.015, $F(1,62) = .08$, $p = .78$, CI = [-0.29 0.22], Fig. 5B). Both slopes were significantly smaller than 0.5 and smaller than the slopes of the non-informative neurons (Supplementary Fig. 7). This indicates that these neurons were influenced more strongly by the reference color, and that the addition of the probe color at load 2 resulted in an unequal interaction. Therefore, DNR was also computationally responsible for a gain of information at load 2, in a specific subset of neurons.

Discussion

Neuronal resources of WM capacity are hemifield independent and gradually allocated.

Our results confirm behavioral findings that have been discussed in detail in an earlier study (*Balakhonov and Rose, 2017*). In brief, we found that the WM capacity of crows is limited to about four items, and that the two visual hemifields are largely independent (i.e., the number of items on one side does not affect change detection performance on the other side). Within each hemifield, performance dropped gradually with the addition of a second and third item but remained above chance. Fittingly, on the neuronal level, we found a markedly reduced amount of color information when the number of colored squares was increased from one to two (roughly 50 % reduction in correct trials). This suggests that WM could be conceptualized as a continuous resource that has to be divided between the two items (*Bays and Husain, 2008; Berg et al., 2012; Wilken and Ma, 2004*), rather than two 'simple' slots that would each have the same amount of information irrespective of the memory load. In contrast, the hemispheric independence we observed would fit a slot-like model, in which the hemispheres as a whole act like discrete slots. A more nuanced version of the slot model ('slots and averaging', (*Zhang and Luck, 2008*)) could also account for graded amounts of information within a limited number of slots (*Fukuda et al., 2010; Zhang and Luck, 2008*), as we found here. The mix of discrete and independent hemispheres with a graded allocation of information between items that we found is comparable to results by (*Buschman et al., 2011*) observed in monkey PFC. On the neuronal level, recurrent connections between neurons within a hemisphere may reduce item differentiation when multiple items are present simultaneously, creating capacity limitations within the hemisphere (*Matsushima and Tanaka, 2014*). A lack of interhemispheric recurrent connections would make processing in the other hemisphere independent. Like in monkeys, WM capacity in crows may therefore result from neuronal activity patterns governed by multiple individual items.

Attentional processes guide WM allocation and maintenance.

One way to circumvent WM failure when item load increases is to allocate attention. Our results suggest that attention may play an important role in crow WM. Capacity limitation became apparent during encoding, as the amount of information at the end of the sample period was affected by the stimulus load. Adding a second and third item to the ipsilateral stimulus array reduced the amount of color information encoded by NCL neurons that carried over into the memory delay. Furthermore, neuronal activity in trials in which the birds made an incorrect response showed only weak encoding during the sample phase without information maintenance

during the memory delay. This fits studies of human WM that have shown attentive filtering during encoding of stimuli influencing WM capacity (*Bays and Husain, 2008; Vogel et al., 2005; Vogel and Machizawa, 2004*), and neuronal correlates of this have been reported for monkeys as well (*Buschman et al., 2011*). Furthermore, attention and WM may be directly linked, as neuronal correlates of WM and attention overlap in PFC neurons (*Lebedev et al., 2004; Panichello and Buschman, 2021*). The independence of hemifields that we observed on the behavioral level (this study and (*Balakhonov and Rose, 2017*)), and found in the neuronal responses could also be related to attention. Adding stimuli in the contralateral hemifield affected neither performance nor information maintained by NCL neurons, whereas additional ipsilateral stimuli strongly reduced both. This fits the influence of attention on WM and hemifield independence, which is consistently accentuated in studies in which attention had to be divided between the two hemifields (*Alvarez and Cavanagh, 2005; Buschman et al., 2011; Cavanagh and Alvarez, 2005; Delvenne, 2005; Delvenne et al., 2011*). Finally, the DNR computation can explain the responses of the neurons that gained information at load 2 through attention. This may appear counter-intuitive and contradictory, considering that the same process is also responsible for the loss of information. However, when attention is overtly directed to a specific (preferred or non-preferred) item within the receptive field of a neuron, the DNR computation shifts its weighting of the normalized response towards the response of the attended item (*Reynolds et al., 1999; Reynolds and Heeger, 2009*). This weighted normalization can produce a difference in the neuronal response to both color identities at load 2, even if the neuronal response was non-informative at load 1. Thus, an attentive process might have enhanced information in WM at higher loads.

As we did not use any form of attentional cueing in our study, we cannot explicitly test such an attention effect. However, we do know that the animals participating in this study can use attentional cues to enhance their WM (*Fongaro and Rose, 2020*). The attention cues used by (*Fongaro and Rose, 2020*) positively affected not only encoding but also the maintenance and retrieval of the information held in WM, comparable to results from monkeys and humans (*Brady and Hampton, 2018; Souza and Oberauer, 2016*). We, therefore, want to emphasize that our data is in line with the interpretation that the birds possibly attended a load 2 stimulus array differently than a load 1 stimulus array in order to enhance their performance in trials with higher loads.

Modern models of mammalian WM capacity are applicable to crows.

Our neuronal recordings offer a mechanistic explanation for the behavioral effects, as we found clear evidence of DNR governing the neuronal responses tied to WM capacity that is in accordance with mammalian models of WM capacity (*Bouchacourt and Buschman, 2019;*

Lundqvist et al., 2016, 2011). The loss of information about color identity (i.e., neuronal response differentiation between colors) can be accounted for by DNR when an item is added to a neuron's receptive field. The normalization of neuronal firing rate diminishes the differentiation between color identities. As such it is analog to neurophysiological responses from visual areas (*Carandini et al., 1997; Reynolds et al., 1999)* and to the prefrontal cortex during spatial WM *(Matsushima and Tanaka, 2014*). The WM model of (*Bouchacourt and Buschman, 2019*) is based solely on data from monkey electrophysiology, and thus implicitly tied to the layered columns of the neocortex. The results we report here show that the model also fits the neurophysiology of WM in crows. However, the picture is incomplete since important aspects of monkeys' WM are still not investigated in crows. Oscillations of local field potentials (LFP) are relevant for how information enters WM and how it is maintained (*Miller et al., 2018*), and have been tied to normalization and competition between items in WM (*Lundqvist et al., 2018*). Thus, the oscillatory interplay of the layers and different regions of the mammalian neocortex are important fields of research to further our understanding of WM. Such aspects are so far completely unknown in crows and their non-layered associative areas. This encourages further investigation into the neuronal circuits of WM in birds.

Therefore, while we cannot, yet, fully equate crow and monkey WM, our results raise two important questions about how WM is implemented on the level of neuronal networks that have implications for our comparative view of crow WM. The first regards the neuronal computations underlying WM. Is there a common canonical computation governing WM, or are there different solutions based on different neuronal architectures? Recent work has shown that the sensory areas of mammals and birds show remarkably similar circuit organization (*Stacho et al., 2020*). However, higher-order associative areas involved in WM, like the LPFC in mammals and the NCL in birds, have distinctly different architectures (*Stacho et al., 2020*). The fact that differently organized areas like LPFC and NCL produce strikingly similar physiological responses, points to shared computational principles. Modeling work already suggests that the competing WM capacity models can be accommodated into a unifying framework based on theoretical neuronal information sampling, where stochastic information sampling (assumed for continuous resource models) can account for item limitations better than fixed information sampling (assumed by the slots and averaging models) (*Schneegans et al., 2020*). Similarly, DNR is already considered to be a general, canonical computation of the nervous system, present in evolutionarily distant phyla, e.g. fruit flies and monkeys (*Carandini and Heeger, 2012*). The second question regards the tradeoff between WM flexibility and capacity (*Bouchacourt and Buschman, 2019*). Is the WM of a crow as flexible as that of a monkey? Our results show that the computations by individual

598 neurons that result in WM capacity limitations are virtually the same in crows and monkeys,
599 highlighting a further aspect of WM that is similar between these animal groups (*Nieder, 2017*).
600 Ultimately, our results were in line with different modern models of WM that implement DNR to
601 explain capacity (*Bouchacourt and Buschman, 2019; Lundqvist et al., 2016, 2011; Schneegans*
602 *et al., 2020*). However, the data we presented cannot carry a definitive conclusion about which of
603 the different models fits best. For example, a tradeoff between flexibility and capacity
604 (*Bouchacourt and Buschman, 2019*) might be present, but further investigation into the models'
605 predictions is required. We do, however, show that mammalian models of WM are in line with WM
606 in birds, which implies that fundamental aspects of WM are shared between these animal groups.

607 **Conclusion.**

608 Together, all these facets of crow WM capacity suggest that the different intricate neuronal
609 architectures that carry out the computations in monkeys and crows have likely been shaped by
610 convergent evolution - into systems that yield similar cognitive performances. The systems may
611 share the same basic mechanisms and thus limitations. Further investigation into the oscillatory
612 dynamics of WM in the avian brain may elucidate if birds also share the prominent limitation of a
613 tradeoff between flexibility and capacity.

614 ## Acknowledgments

615 We would like to thank Mikael Lundqvist for helpful comments on an earlier version of the
616 manuscript. This work was supported by a Volkswagen Foundation Freigeist Fellowship
617 awarded to J.R.

618 ## Competing interests

619 The authors declare no competing interests.

620 ## Author contributions

621 L.A.H. analyzed the data, performed spike sorting and wrote the manuscript. D.B. performed
622 animal training and data acquisition. E.F. performed spike sorting. A.N. provided methodological
623 and infrastructural support and edited the manuscript. J.R. conceptualized, planned, and
624 supervised the experiment, provided funding, and edited the manuscript.

625

References

Alvarez, G.A., Cavanagh, P., 2005. Independent Resources for Attentional Tracking in the Left and Right Visual Hemifields. Psychol. Sci. 16, 637–643. https://doi.org/10.1111/j.1467-9280.2005.01587.x

Awh, E., Barton, B., Vogel, E.K., 2007. Visual Working Memory Represents a Fixed Number of Items Regardless of Complexity. Psychol. Sci. 18, 622–628. https://doi.org/10.1111/j.1467-9280.2007.01949.x

Balakhonov, D., Rose, J., 2017. Crows Rival Monkeys in Cognitive Capacity. Sci. Rep. 7, 8809. https://doi.org/10.1038/s41598-017-09400-0

Bays, P.M., Husain, M., 2008. Dynamic Shifts of Limited Working Memory Resources in Human Vision. Science 321, 851–854. https://doi.org/10.1126/science.1158023

Berg, R. van den, Shin, H., Chou, W.-C., George, R., Ma, W.J., 2012. Variability in encoding precision accounts for visual short-term memory limitations. Proc. Natl. Acad. Sci. 109, 8780–8785. https://doi.org/10.1073/pnas.1117465109

Bouchacourt, F., Buschman, T.J., 2019. A Flexible Model of Working Memory. Neuron 103, 147-160.e8. https://doi.org/10.1016/j.neuron.2019.04.020

Brady, R.J., Hampton, R.R., 2018. Post-encoding control of working memory enhances processing of relevant information in rhesus monkeys (Macaca mulatta). Cognition 175, 26–35. https://doi.org/10.1016/j.cognition.2018.02.012

Brainard, D.H., 1997. The psychophysics toolbox. Spat. Vis. 10, 433–436. https://doi.org/10.1163/156856897x00357

Buschman, T.J., Siegel, M., Roy, J.E., Miller, E.K., 2011. Neural substrates of cognitive capacity limitations. Proc. Natl. Acad. Sci. 108, 11252–11255. https://doi.org/10.1073/pnas.1104666108

Carandini, M., Heeger, D.J., 2012. Normalization as a canonical neural computation. Nat. Rev. Neurosci. 13, 51–62. https://doi.org/10.1038/nrn3136

Carandini, M., Heeger, D.J., Movshon, J.A., 1997. Linearity and Normalization in Simple Cells of the Macaque Primary Visual Cortex. J. Neurosci. 17, 8621–8644. https://doi.org/10.1523/JNEUROSCI.17-21-08621.1997

Cavanagh, P., Alvarez, G.A., 2005. Tracking multiple targets with multifocal attention. Trends Cogn. Sci. 9, 349–354. https://doi.org/10.1016/j.tics.2005.05.009

Cowan, N., 2017. The many faces of working memory and short-term storage. Psychon. Bull. Rev. 24, 1158–1170. https://doi.org/10.3758/s13423-016-1191-6

Cowan, N., 2001. The magical number 4 in short-term memory: A reconsideration of mental storage capacity. Behav. Brain Sci. 24, 87–114. https://doi.org/10.1017/S0140525X01003922

Delvenne, J.-F., 2005. The capacity of visual short-term memory within and between hemifields. Cognition 96, B79–B88. https://doi.org/10.1016/j.cognition.2004.12.007

Delvenne, J.-F., Kaddour, L.A., Castronovo, J., 2011. An electrophysiological measure of visual short-term memory capacity within and across hemifields. Psychophysiology 48, 333–336. https://doi.org/10.1111/j.1469-8986.2010.01079.x

Diekamp, B., Gagliardo, A., Güntürkün, O., 2002. Nonspatial and subdivision-specific working memory deficits after selective lesions of the avian prefrontal cortex. J. Neurosci. 22, 9573–9580.

Fongaro, E., Rose, J., 2020. Crows control working memory before and after stimulus encoding. Sci. Rep. 10, 1–10. https://doi.org/10.1038/s41598-020-59975-4

Fukuda, K., Awh, E., Vogel, E.K., 2010. Discrete capacity limits in visual working memory. Curr. Opin. Neurobiol., Cognitive neuroscience 20, 177–182. https://doi.org/10.1016/j.conb.2010.03.005

675 Güntürkün, O., Bugnyar, T., 2016. Cognition without Cortex. Trends Cogn. Sci. 20, 291–303.
676 https://doi.org/10.1016/j.tics.2016.02.001
677 Hartmann, K., Veit, L., Nieder, A., 2018. Neurons in the crow nidopallium caudolaterale encode
678 varying durations of visual working memory periods. Exp. Brain Res. 236, 215–226.
679 https://doi.org/10.1007/s00221-017-5120-3
680 Heeger, D.J., 1992. Normalization of cell responses in cat striate cortex. Vis. Neurosci. 9, 181–
681 197. https://doi.org/10.1017/S0952523800009640
682 Husband, S., Shimizu, T., 2001. Evolution of the Avian Visual System, in: Avian Visual
683 Cognition [On-Line]. Available: Pigeon.Psy.Tufts.Edu/Avc/Husband/.
684 Johnson, M.K., McMahon, R.P., Robinson, B.M., Harvey, A.N., Hahn, B., Leonard, C.J., Luck,
685 S.J., Gold, J.M., 2013. The relationship between working memory capacity and broad
686 measures of cognitive ability in healthy adults and people with schizophrenia.
687 Neuropsychology 27, 220–229. https://doi.org/10.1037/a0032060
688 Kornblith, S., Buschman, T.J., Miller, E.K., 2016. Stimulus Load and Oscillatory Activity in
689 Higher Cortex. Cereb. Cortex 26, 3772–3784. https://doi.org/10.1093/cercor/bhv182
690 Kröner, S., Güntürkün, O., 1999. Afferent and efferent connections of the caudolateral
691 neostriatum in the pigeon (Columba livia): A retro- and anterograde pathway tracing
692 study. J. Comp. Neurol. 407, 228–260. https://doi.org/10.1002/(sici)1096-
693 9861(19990503)407:2<228::aid-cne6>3.0.co;2-2
694 Lebedev, M.A., Messinger, A., Kralik, J.D., Wise, S.P., 2004. Representation of Attended
695 Versus Remembered Locations in Prefrontal Cortex. PLOS Biol. 2, e365.
696 https://doi.org/10.1371/journal.pbio.0020365
697 Luck, S.J., Vogel, E.K., 1997. The capacity of visual working memory for features and
698 conjunctions. Nature 390, 279. https://doi.org/10.1038/36846
699 Lundqvist, M., Herman, P., Lansner, A., 2011. Theta and Gamma Power Increases and
700 Alpha/Beta Power Decreases with Memory Load in an Attractor Network Model. J. Cogn.
701 Neurosci. 23, 3008–3020. https://doi.org/10.1162/jocn_a_00029
702 Lundqvist, M., Herman, P., Warden, M.R., Brincat, S.L., Miller, E.K., 2018. Gamma and beta
703 bursts during working memory readout suggest roles in its volitional control. Nat.
704 Commun. 9, 394. https://doi.org/10.1038/s41467-017-02791-8
705 Lundqvist, M., Rose, J., Herman, P., Brincat, S.L., Buschman, T.J., Miller, E.K., 2016. Gamma
706 and Beta Bursts Underlie Working Memory. Neuron 90, 152–164.
707 https://doi.org/10.1016/j.neuron.2016.02.028
708 Ma, W.J., Husain, M., Bays, P.M., 2014. Changing concepts of working memory. Nat. Neurosci.
709 17, 347–356. https://doi.org/10.1038/nn.3655
710 Matsushima, A., Tanaka, M., 2014. Different Neuronal Computations of Spatial Working
711 Memory for Multiple Locations within versus across Visual Hemifields. J. Neurosci. 34,
712 5621–5626. https://doi.org/10.1523/JNEUROSCI.0295-14.2014
713 Miller, E.K., Lundqvist, M., Bastos, A.M., 2018. Working Memory 2.0. Neuron 100, 463–475.
714 https://doi.org/10.1016/j.neuron.2018.09.023
715 Nieder, A., 2017. Inside the corvid brain—probing the physiology of cognition in crows. Curr.
716 Opin. Behav. Sci., Comparative cognition 16, 8–14.
717 https://doi.org/10.1016/j.cobeha.2017.02.005
718 Oberauer, K., Lewandowsky, S., Awh, E., Brown, G.D.A., Conway, A., Cowan, N., Donkin, C.,
719 Farrell, S., Hitch, G.J., Hurlstone, M.J., Ma, W.J., Morey, C.C., Nee, D.E., Schweppe, J.,
720 Vergauwe, E., Ward, G., 2018. Benchmarks for models of short-term and working
721 memory. Psychol. Bull. 144, 885–958. https://doi.org/10.1037/bul0000153
722 Panichello, M.F., Buschman, T.J., 2021. Shared mechanisms underlie the control of working
723 memory and attention. Nature 1–5. https://doi.org/10.1038/s41586-021-03390-w

Reynolds, J.H., Chelazzi, L., Desimone, R., 1999. Competitive Mechanisms Subserve Attention in Macaque Areas V2 and V4. J. Neurosci. 19, 1736–1753. https://doi.org/10.1523/JNEUROSCI.19-05-01736.1999

Reynolds, J.H., Heeger, D.J., 2009. The Normalization Model of Attention. Neuron 61, 168–185. https://doi.org/10.1016/j.neuron.2009.01.002

Rinnert, P., Kirschhock, M.E., Nieder, A., 2019. Neuronal Correlates of Spatial Working Memory in the Endbrain of Crows. Curr. Biol. 29, 2616-2624.e4. https://doi.org/10.1016/j.cub.2019.06.060

Rose, J., Colombo, M., 2005. Neural correlates of executive control in the avian brain. Plos Biol. 3, 1139–1146. https://doi.org/10.1371/journal.pbio.0030190

Rose, J., Otto, T., Dittrich, L., 2008. The Biopsychology-Toolbox: A free, open-source Matlab-toolbox for the control of behavioral experiments. J. Neurosci. Methods 175, 104–107. https://doi.org/10.1016/j.jneumeth.2008.08.006

Rossant, C., Kadir, S.N., Goodman, D.F.M., Schulman, J., Hunter, M.L.D., Saleem, A.B., Grosmark, A., Belluscio, M., Denfield, G.H., Ecker, A.S., Tolias, A.S., Solomon, S., Buzsaki, G., Carandini, M., Harris, K.D., 2016. Spike sorting for large, dense electrode arrays. Nat Neurosci 19, 634–641. https://doi.org/10.1038/nn.4268

Schneegans, S., Taylor, R., Bays, P.M., 2020. Stochastic sampling provides a unifying account of visual working memory limits. Proc. Natl. Acad. Sci. 117, 20959–20968. https://doi.org/10.1073/pnas.2004306117

Souza, A.S., Oberauer, K., 2016. In search of the focus of attention in working memory: 13 years of the retro-cue effect. Atten. Percept. Psychophys. 78, 1839–1860. https://doi.org/10.3758/s13414-016-1108-5

Stacho, M., Herold, C., Rook, N., Wagner, H., Axer, M., Amunts, K., Güntürkün, O., 2020. A cortex-like canonical circuit in the avian forebrain. Science 369. https://doi.org/10.1126/science.abc5534

Troscianko, J., von Bayern, A.M.P., Chappell, J., Rutz, C., Martin, G.R., 2012. Extreme binocular vision and a straight bill facilitate tool use in New Caledonian crows. Nat. Commun. 3, 1110. https://doi.org/10.1038/ncomms2111

Veit, L., Hartmann, K., Nieder, A., 2014. Neuronal Correlates of Visual Working Memory in the Corvid Endbrain. J. Neurosci. 34, 7778–7786. https://doi.org/10.1523/jneurosci.0612-14.2014

Veit, L., Nieder, A., 2013. Abstract rule neurons in the endbrain support intelligent behaviour in corvid songbirds. Nat. Commun. 4, 11. https://doi.org/10.1038/ncomms3878

Vogel, E.K., Machizawa, M.G., 2004. Neural activity predicts individual differences in visual working memory capacity. Nature 428, 748–751. https://doi.org/10.1038/nature02447

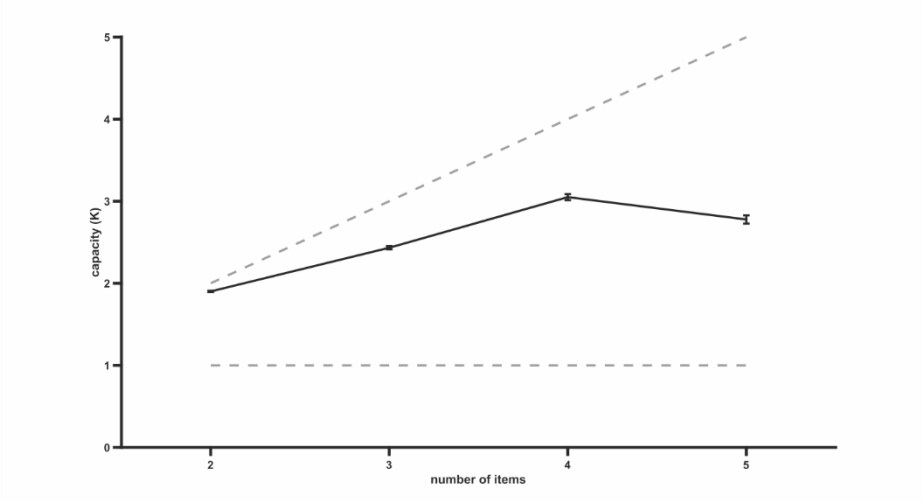
Vogel, E.K., McCollough, A.W., Machizawa, M.G., 2005. Neural measures reveal individual differences in controlling access to working memory. Nature 438, 500–503. https://doi.org/10.1038/nature04171

Waldmann, C., Güntürkün, O., 1993. The dopaminergic innervation of the pigeon caudolateral forebrain - immunocytochemical evidence for a prefrontal cortex in birds. Brain Res. 600, 225–234. https://doi.org/10.1016/0006-8993(93)91377-5

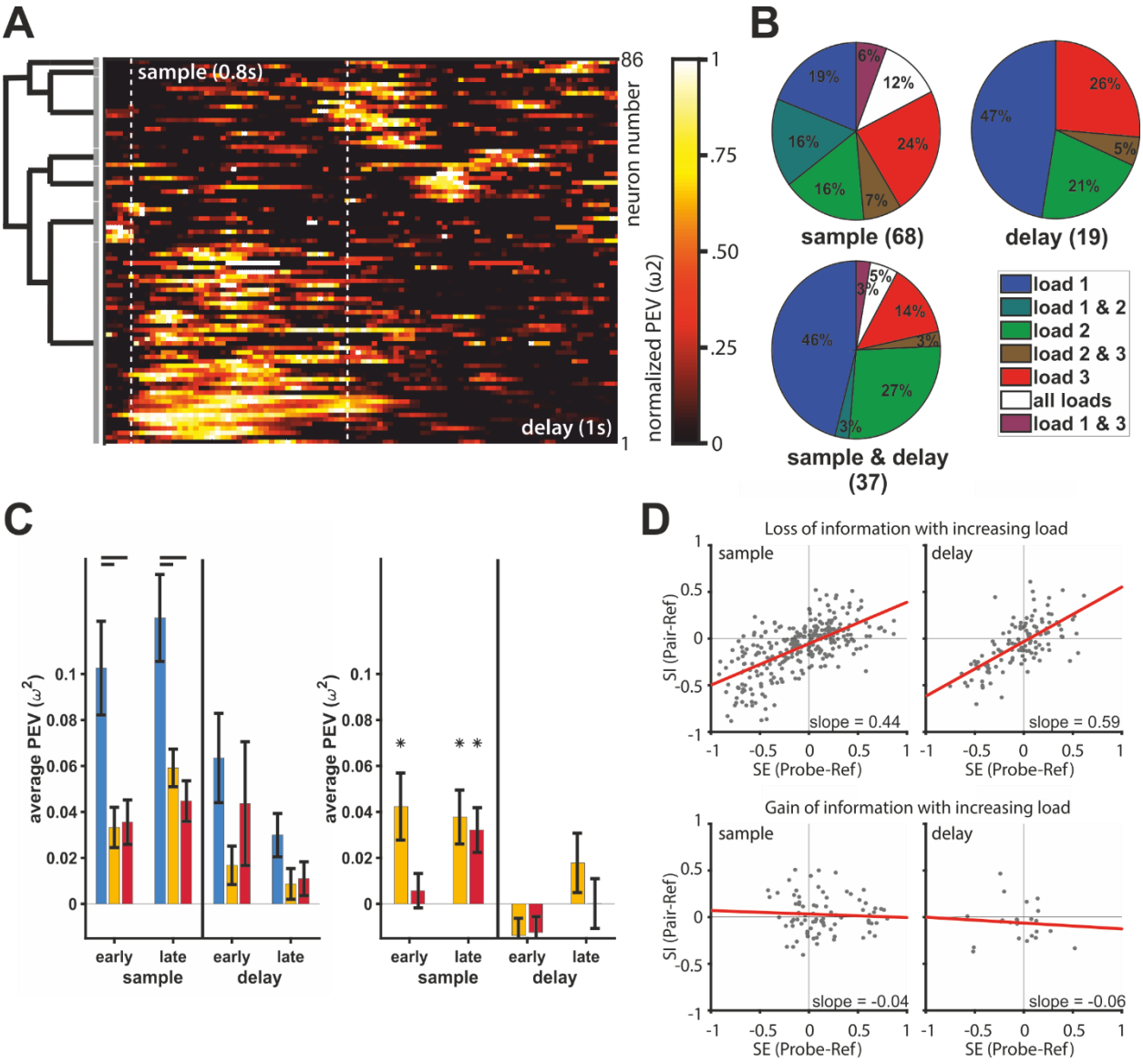
Wilken, P., Ma, W.J., 2004. A detection theory account of change detection. J. Vis. 4, 11–11. https://doi.org/10.1167/4.12.11

Zhang, W., Luck, S.J., 2008. Discrete fixed-resolution representations in visual working memory. Nature 453, 233–235. https://doi.org/10.1038/nature06860

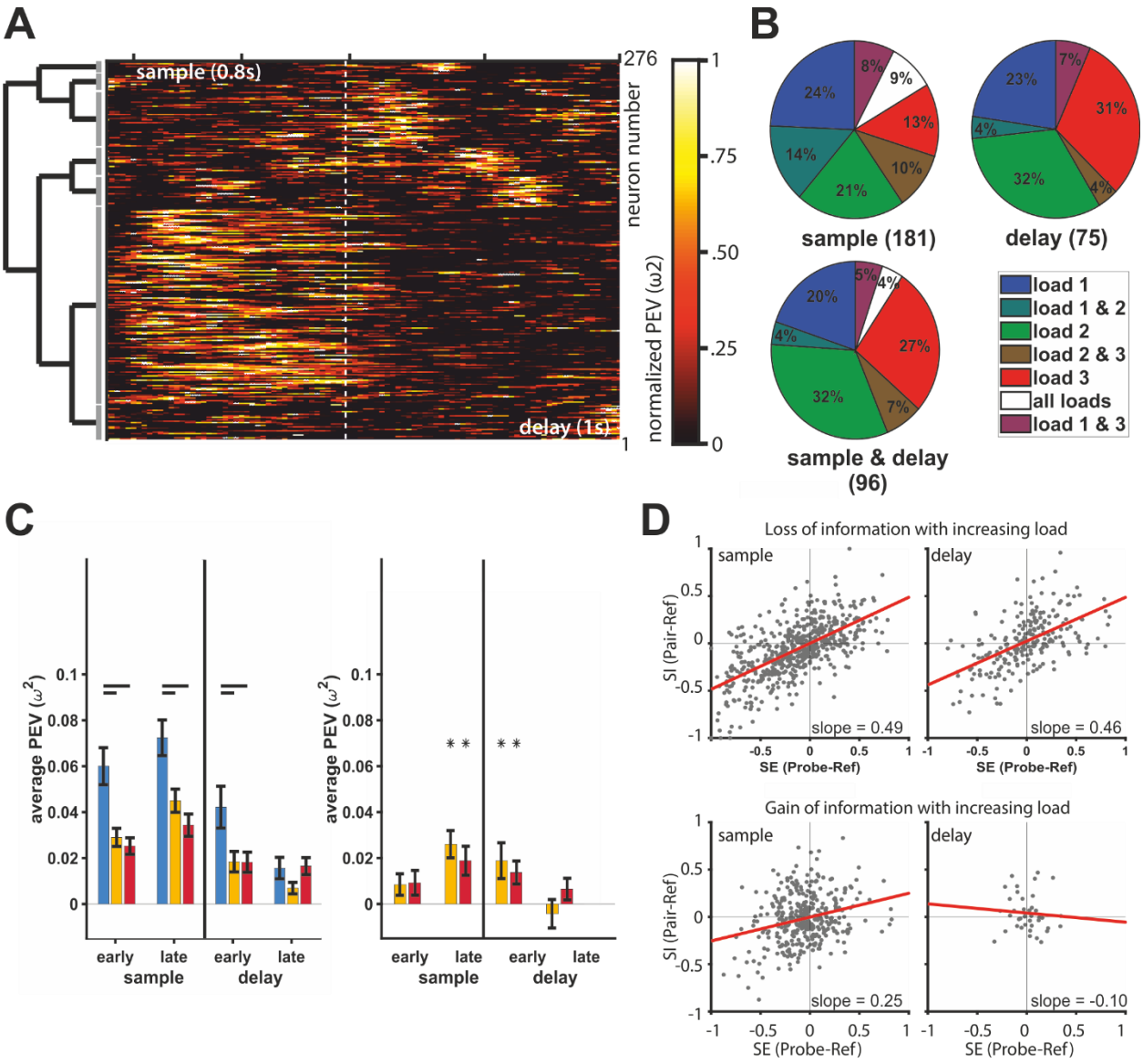
Supplementary Material



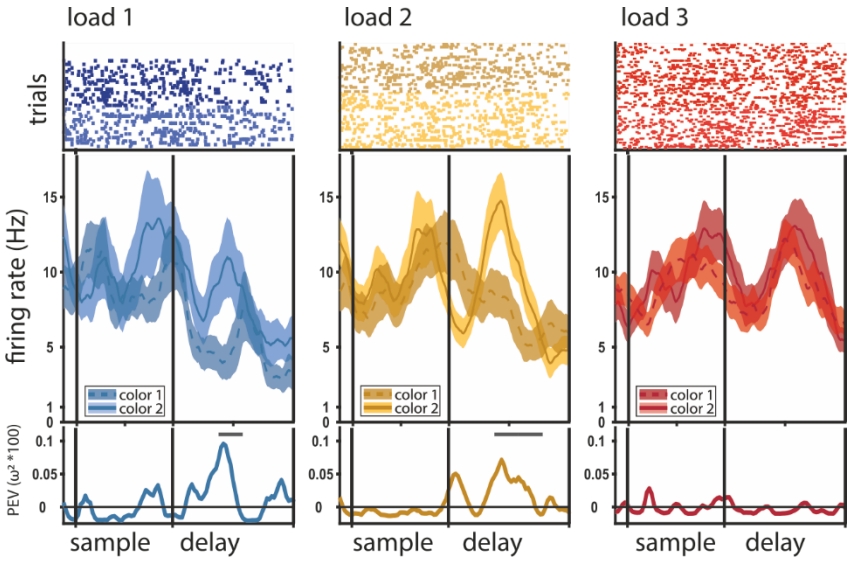
Supplementary Figure 1: Capacity of crow WM. Line indicates capacity K at different loads. The peak at 4 items indicates the capacity. Dashed lines indicate maximum capacity and fixed capacity of 1. Error bars indicate the standard error of the mean.



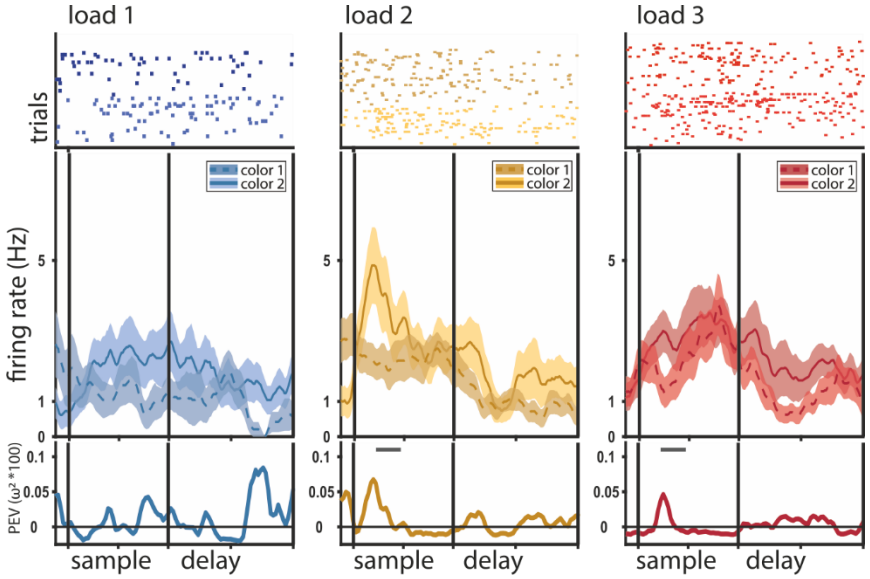
Supplementary Figure 2: Overview of analyses for bird 1. (A) The neuronal population can be best described by 7 individual clusters. (B) Percentages (rounded) of significant neurons in individual load conditions for sample (n = 68), delay (n = 19), and sample & delay (n = 37). (C) On correct trials (left) color is represented during the early and late phase of the sample and, to a lesser degree, during the early and late delay. On error trials (right), color information can be found in the early sample phase at load 2, and in the late sample phase at loads 2 & 3 (asterisks). Analysis of load 1 error trials was omitted due to their very low abundance. Statistical comparisons of correct vs. error trial information were performed on sub-sampled correct trials. Early and late sample each 400 ms, early and late delay each 500 ms, error bars indicate the standard error of the mean. (D) Divisive normalization-like regularization was observable for neuronal responses of neurons losing information (top) but not for neurons gaining color information at load 2 (bottom). Selectivity (SE) indicates how much the neuronal response is influenced by a color, relative to a second color when either is presented alone. Sensory interaction (SI) indicates how much the neuronal response is influenced by either color when both were displayed simultaneously. Slopes close to 0.5 indicate an equal influence of both colors. Slopes < 0.5, or > 0.5 indicate a weighted influence of a color. (Top) Information carrying neurons in the sample (n = 35; left) and delay (n = 15; right) population. Bottom) Information gaining neurons in the sample (n = 10; left) and delay (n = 3; right) population. The red line indicates the regression fit.



Supplementary Figure 3: Overview of analyses for bird 2. (A) The neuronal population can be best described by 7 individual clusters. (B) Percentages (rounded) of significant neurons in individual load conditions for sample (n = 181), delay (n = 75), and sample & delay (n = 96). (C) On correct trials (left) color is represented during the early and late phase of the sample and, to a lesser degree, during the early and late delay. On error trials (right), color information can be found in the late sample phase at loads 2 & 3 (asterisks). Analysis of load 1 error trials was omitted due to their very low abundance. Statistical comparisons of correct vs. error trial information were performed on sub-sampled correct trials. Early and late sample each 400 ms, early and late delay each 500 ms, error bars indicate the standard error of the mean. (D) Divisive normalization-like regularization was observable for neuronal responses of neurons losing information (top) but not for neurons gaining color information at load 2 (bottom). Selectivity (SE) indicates how much the neuronal response is influenced by a color, relative to a second color when either is presented alone. Sensory interaction (SI) indicates how much the neuronal response is influenced by either color when both were displayed simultaneously. Slopes close to 0.5 indicate an equal influence of both colors. Slopes < 0.5, or > 0.5 indicate a weighted influence of a color. (Top) Information carrying neurons in the sample (n = 70; left) and delay (n = 28; right) population. Bottom) Information gaining neurons in the sample (n = 46; left) and delay (n = 5; right) population. The red line indicates the regression fit.

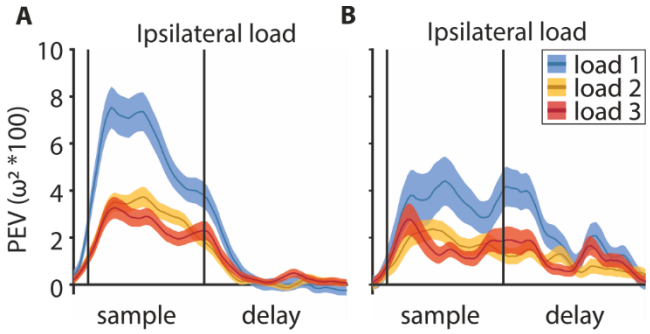


Supplementary Figure 4: Color discrimination in the neuronal response (information, PEV) decreases with load. Example of a delay neuron with color information decline, at load 1 (blue), load 2 (green), and load 3 (red). Top: raster plot, where every dot represents a single spike during the individual trials (rows of dots); middle: peri-stimulus-time histogram (PSTH) of average firing rate (solid line for color ID 1, dashed line for color ID 2) with the standard error of the mean (shaded areas); bottom: percent explained variance of color identity (a measure of information about color) along the trial, the line at the top of the y-axis indicates significant bins.



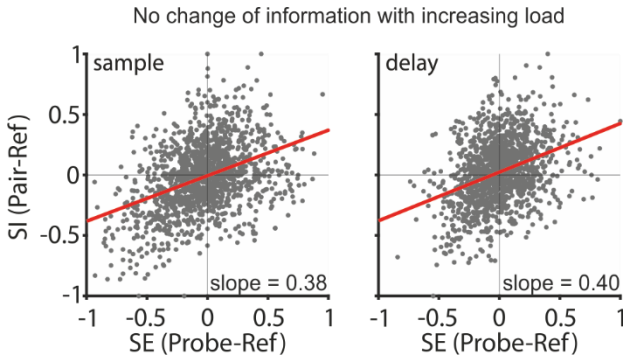
Supplementary Figure 5: Color discrimination in the neuronal response (information, PEV) increases with load. Example of a sample neuron with color information gain; at load 1 (blue), at load 2 (green), and load 3 (red). Top: raster plot, where every dot represents a single spike during the individual trials (rows of dots); middle: peri-stimulus-time histogram (PSTH) of average firing rate (solid line for color ID 1, dashed line for color ID 2) with the standard error of the mean (shaded areas); bottom: percent explained variance of color identity (a measure of information about color) along the trial, the line at the top of the y-axis indicates significant bins.

Notably, the 'delay-populations' also showed an elevated level of information during the sample, whereas the 'sample-populations' did not show an elevated level of information during the delay (Supplementary Fig. 4).



Supplementary Figure 6: Sample population (A) and delay population (B), same as Fig. 4A with full time axis.

Non-informative sample neurons had a fitted slope of 0.38 (R^2_{adj} = .16, F(1,1366) = 258.08, p < .001), significantly smaller than 0.5 (*CI* = [0.33 0.42]). Delay neurons had a slope of 0.40, also significantly smaller than 0.5 (R^2_{adj} = .13, F(1,1366) = 197.51, p < .001, *CI* = [0.35 0.46]). This indicates that for these neurons the reference color influenced firing rate more than the probe color. This smaller slope is not related to the amount of information encoded for the individual colors (which determined the favorite location). It does however indicate those non-informative neurons were influenced by any color at their favorite location and thereby might have been informative about if the favorite location had a color but not about what color.



Supplementary Figure 7: Divisive normalization-like regularization was observable for neuronal responses of neurons without significant information. Both phases contain the same neurons (n = 171). Selectivity (SE) indicates how much the neuronal response is influenced by a color, relative to a second color when either is presented alone. Sensory interaction (SI) indicates how much the neuronal response is influenced by either color when both were displayed simultaneously. Slopes close to 0.5 indicate an equal influence of both colors. Slopes < 0.5, or > 0.5 indicate a weighted influence of a color. The red line indicates the regression fit.