bioRxiv preprint doi: https://doi.org/10.1101/076554; this version posted September 21, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Reading positional codes with fMRI: Problems and solutions

Kristjan Kalm,* Dennis Norris

Cognition and Brain Sciences Unit, Medical Research Council
15 Chaucer Road, Cambridge, CB2 7EF, UK

* E-mail: kristjan.kalm@mrc-cbu.cam.ac.uk.

September 21, 2016

Neural mechanisms which bind items into sequences have been investigated in a large body of research in animal neurophysiology and human neuroimaging. However, a major problem in interpreting this data arises from a fact that several unrelated processes, such as memory load, sensory adaptation, and reward expectation, also change in a consistent manner as the sequence unfolds. In this paper we show that the problem of extracting neural data about the structure of a sequence is especially acute for fMRI, which is almost exclusively the modality used in human experiments. We show that such fMRI results must be treated with caution and in many cases the assumed neural representation might actually reflect unrelated processes.

Introduction

One of the most important features of human cognition is the ability to bind individual events into a sequence. Almost any complex task requires us to remember not only the individual elements but also the order in which they occurred. For example, two tasks such as starting a car and stopping it might share the same events but in different order. All computational models of sequence processing acknowledge this distinction between the representations of items

bioRxiv preprint doi: https://doi.org/10.1101/076554; this version posted September 21, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

in memory and the representation of the order in which they occur (Henson & Burgess, 1997; Page & Norris, 1998). The view that item's position in the sequence is encoded separately and independently of their identity has been also suggested by decades of research in human behaviour and animal neurophysiology.

Neurons in the monkey prefrontal cortex (PFC) have been found to be selective for each position in a learned sequence (Nakajima, Hosaka, Mushiake, & Tanji, 2009; Averbeck, Crowe, Chafee, & Georgopoulos, 2003; Inoue & Mikami, 2006; Naya & Suzuki, 2011). Figure 1C gives an example of a simple positional code showing the responses of position-sensitive neurons from monkey supplementary motor area as observed by Berdyyeva and Olson (2010). Other research on animal neurophysiology has suggested that the hippocampus encodes the position of items in a sequence (Mankin et al., 2012; Manns, Howard, & Eichenbaum, 2007; Naya & Suzuki, 2011; Pastalkova, Itskov, Amarasingham, & Buzsáki, 2008), with some authors proposing the existence of 'time cells' tracking the temporal position of items in a sequence (MacDonald, Lepage, Eden, & Eichenbaum, 2011; MacDonald, Carrow, Place, & Eichenbaum, 2013). From hereon we refer to such neural representation of the item's position in the sequence as *positional code*. The extensive literature on the neural representation of the positional code is summarised in Table 1.

However, interpreting a neural signal tracking the positional code suffers from a major methodological problem: items in different positions necessarily differ along other dimensions too. For example, in a memory task, memory load will be greater at position three than position two. Changes in neural activity that are sensitive to memory load might therefore give the appearance of coding position. An item in position n will always be associated with a load of n items. Any neural index of load will therefore consistently be in a different state for items in different positions. An item in position n also occurs at a later time than item $n-1$. Sensory adaptation might change the neural response to items as the sequence progresses. Such a signal could also masquerade as a positional code. Any or all of these factors might therefore lead to a differential neural response which would correlate with the position of an item in a sequence, but which might play no role in determining how the brain codes temporal position.

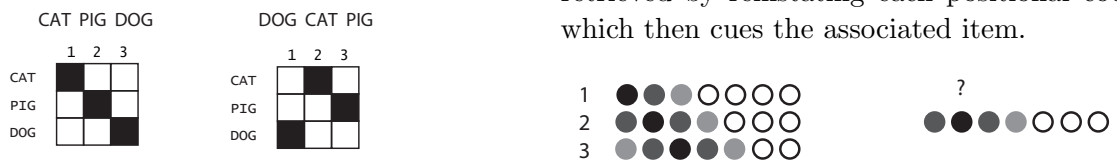
bioRxiv preprint doi: https://doi.org/10.1101/076554; this version posted September 21, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

In their analysis of how we can measure information in the brain, De-Wit, Alexander, Ekroll, and Wagemans (2016) made a contrast between "cortex as receiver" and "experimenter as receiver". There may be ways in which we as experimenters can decode neural states to recover information about temporal position, but what we would like to do is to identify specifically those neural representations that the cortex uses to represent position and to drive behaviour.

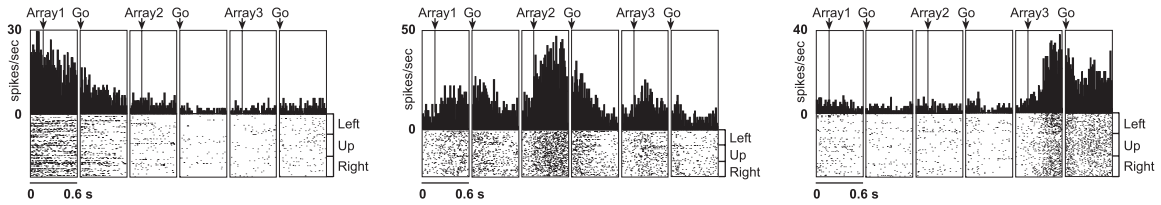
Figure 1: Sequence representation and temporal position

(A) Representation of two sequences as mappings between item codes and temporal position codes.

(B) Left: representation of temporal position in a 7-item sequence. The variance around positional signal is coded in terms of the darkness of the circle. Right: the order position is retrieved by reinstating each positional code which then cues the associated item.



(C) Examples of temporal position selective neurons from Berdyyeva and Olson (2010). From left to right: pre-supplementary motor area neuron selective for 1st position, supplementary eye field neuron selective for 2nd position, and supplementary motor area neuron selective for the 3rd position in the serial object task.



In this paper we show that dissociating positional 'read-out' from a dedicated positional code is an especially difficult problem for fMRI data. We show that fMRI data acquired from sequentially presented stimuli suffer from several confounds. First, we show that with any sequence processing task there are experimental variables which are collinear with the positional signal (e.g. memory load, sensory adaptation, etc.) and which can serve as a positional code. Second, we show how interference between stimulus representations, task phases, and measurement modalities can also lead to a similar positional read-outs indistinguishable from a dedicated positional code. Importantly these correlated effects do not simply result in a uni-

bioRxiv preprint doi: https://doi.org/10.1101/076554; this version posted September 21, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

variate change in signal that varies across position but also change the pattern of information that can be read out by multivariate methods.

The problem of interpreting a positional signal is especially relevant since neural data on human sequence processing comes almost exclusively from fMRI studies (Table 1). Our simulations and experimental data show that results from fMRI experiments studying the positional code must be treated with caution. Specifically, in many cases the assumed positional code might actually reflect processes which are correlated with position in the sequence instead.

Positional code from collinear processes

Any signal tracking the position of an item in a sequence will be collinear with a number of cognitive processes:

- Memory load – signal for position n will always co-occur with a memory load of n items when storing a sequence. Any neural index of load will therefore always reflect the progression of sequence.
- Sensory adaptation – neural responses in the human sensory cortex have been shown to monotonically decrease as a response to sequentially presented stimuli (Henson & Rugg, 2003; Summerfield, Trittschuh, Monti, Mesulam, & Egner, 2008; Larsson & Smith, 2012). Any signal that monotonically changes over sequence positions can be used to read out position-like code.
- Reward – in most animal studies the subject is rewarded after successfully attending or recalling a sequence. This means that the next item in a sequence is always closer to the reward. Neurons tracking the temporal proximity of reward have been described in both monkey and rodent studies (Berdyyeva & Olson, 2011; MacDonald et al., 2011; Naya & Suzuki, 2011).
- Passage of time – signal for position n always occurs after the signal for position $n-1$.

All these processes represent a change in the cognitive state of the participant throughout the processed sequence, and hence will necessarily be collinear with any positional code. It

bioRxiv preprint doi: https://doi.org/10.1101/076554; this version posted September 21, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

follows that in the analysis of experiments on temporal order it is necessary to distinguish between a dedicated positional code and a positional read-out from collinear processes.

Next, we provide two examples of positional read-out based on human fMRI data. In the first example we show how sensory adaptation in the sensory cortices can be interpreted as a positional signal. In the second example we show how differences in retinotopic activation over the course of the sequence can similarly provide positional read-out. In the final part of the section we provide simulations which explore whether it is possible to develop methods to subtract the effects of such collinear processes from sequentially obtained data.

Sensory adaptation

Sensory adaptation across sequence positions has been observed in a number of fMRI studies of sequence processing as a decreasing univariate signal over positions (Summerfield et al., 2008; Larsson & Smith, 2012; Henson & Rugg, 2003). Note that an inverse trend, where the univariate signal increases over sequence positions, has also been observed (Kalm & Norris, 2016). The latter most likely reflects the attenuation of the BOLD signal in response to sequentially presented stimuli as reported in other fMRI studies on human STM (Rottschy et al., 2012; Wager & Smith, 2003). However, the direction of the univariate change is unimportant as any consistent change over sequence positions will permit position decoding.

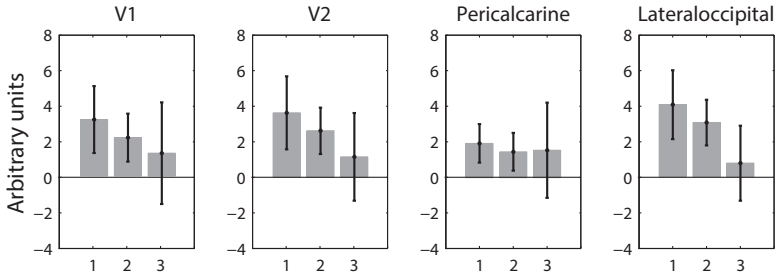
Here we used two human fMRI datasets obtained with a sequence processing task (Kalm & Norris, 2014, 2016) to carry out a classification analysis of item position in a sequence. In both cases we chose the sensory cortex of the presented stimuli as a region of interest (ROI): in the first experiment the sequences were presented auditorily (Kalm & Norris, 2014) and in the second visually (Kalm & Norris, 2016). Since in both experiments sensory areas responded differentially to sequence positions (Figure 2A) linear classification analysis can be used to predict the position of the items significantly above chance (Figure 2B). However, in both cases the signal changes were uniform across all voxels in the brain region suggesting not a dedicated positional code, but sensory adaptation or change in measurement noise. Sensory adaptation thus serves as a clear example how a monotonically changing signal can be read out by an

bioRxiv preprint doi: https://doi.org/10.1101/076554; this version posted September 21, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

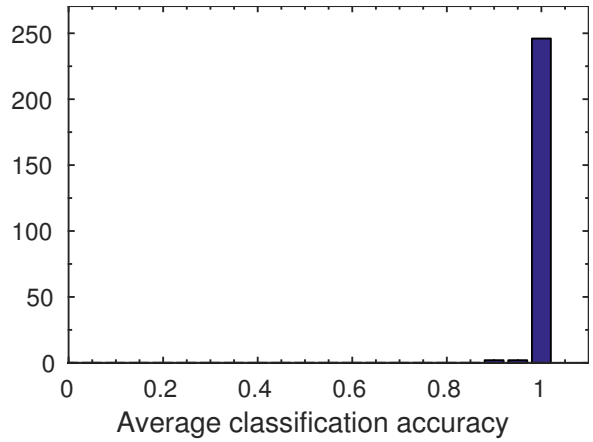
experimenter as a positional code.

Figure 2: Sensory adaptation in the sensory cortex and decoding order position.

(A) Uniform signal change over 3-item sequences in sensory brain areas as reported in Kalm and Norris (2016).



(B) Distribution of average linear classification accuracy values of item position in a sequence in a single subject V1.



Retinotopic activation

In the example above (Sensory adaptation) the population of neural units (sensory cortex) responded uniformly to sequentially presented stimuli. Next we present a case where neural units *within* the population respond differentially across the sequence. We use fMRI data from a visual sequence processing task to show that the response in the primary visual cortex can be used to predict the position of the item in the sequence. However, this is possible not because of any positional code but task-selective voxels in the visual cortex.

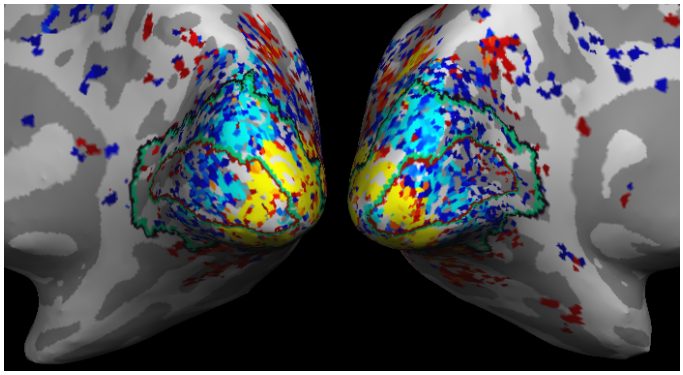
In Kalm and Norris (2016) participants had to attend a sequence of visually presented images followed by a manual response indicating the order of the items (Figure 16). Importantly, all

bioRxiv preprint doi: https://doi.org/10.1101/076554; this version posted September 21, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

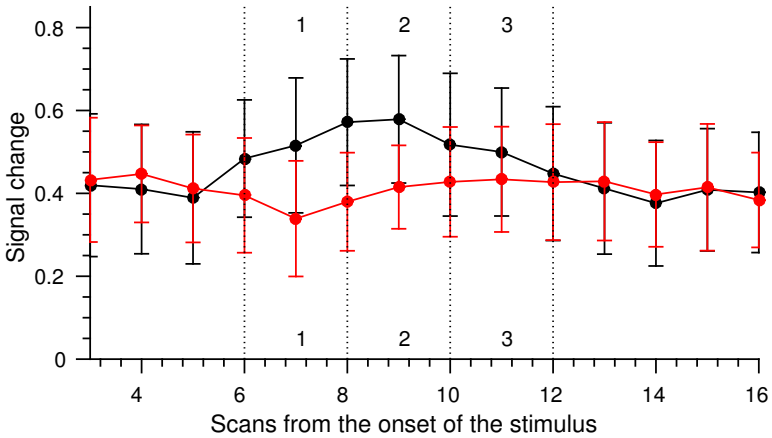
images were controlled for luminance and cropped to ensure that each image appeared in a similar retinal area: all stimuli subtended a 6° visual angle around the fixation point in order to elicit an approximately foveal retinotopic representation. As a result, all sequence items elicited approximately similar retinotopic response in the foveal area of the primary visual cortex (Figure 3A).

Figure 3: Interference between task phases: retinotopic suppression

(A) Activation and suppression in V1 averaged across all stimuli for a single participants from the study. The activated voxels (yellow, $p < 0.001$) mark the foveal part of the visual cortex driven by our stimuli (presented at 6° visual angle).



(B) Peristimulus time histogram of sequence presentation of two groups of voxels from a single participant's V1. The black line marks the average of the voxels activated by the stimuli and the red line marks the average of the voxels suppressed by the stimuli. Dashed vertical lines indicate the time bins where sequence items were presented.



The authors observed that the activation of the retinotopically driven voxels was correlated

bioRxiv preprint doi: https://doi.org/10.1101/076554; this version posted September 21, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

with the relative suppression of the voxels outside of the retinotopically activated areas (Figure 3A). Such suppression has been observed as a function of stimulus location in the visual field (Maier et al., 2008) and attention (Gouws et al., 2014; Smith, Singh, & Greenlee, 2000). Importantly, the amount of activation and suppression changed across sequence positions. Since the sequence items were presented in immediate succession, the extent of retinotopic suppression and activation varied as a function of item's position in the sequence. Figure 3B shows data from a single participant's V1, where voxels are split into two groups: retinotopically activated (red-yellow on Figure 3A) and suppressed (blue-cyan on Figure 3A) represented by red and blue lines. As the activation and suppression of two different sets of voxels changes across positions, a linear classification algorithm can use the difference between activated and suppressed voxels, or the difference between the red and blue lines on Figure 3B, to reliably predict the item's position.

This can be further illustrated when linear discriminant analysis (LDA) class boundaries based on item position are plotted with following sets of voxels from V1:

1. All voxels (including both retinotopically activated and suppressed voxels)
2. Only activated voxels ($p < 0.01$)
3. Only suppressed voxels ($p < 0.01$)

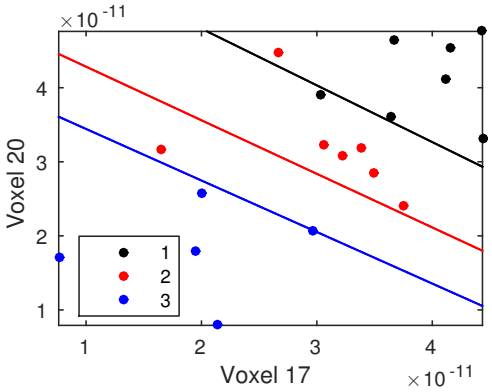
LDA shows that the linear classifier is only able to reliably predict the position of the item when both activated and suppressed voxels in the brain region are included (Figure 4, top row). The classification is at chance level if only one set of voxels are used (Figure 4, row 2-3).

Next we carry out a simulation of sequentially generated fMRI data to explore whether both uniformly and differentially proceeding collinear processes could be controlled for when trying to extract a positional code.

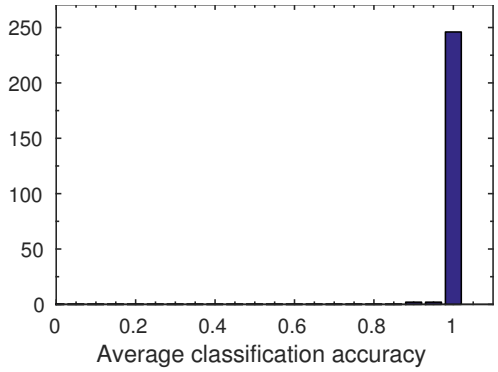
bioRxiv preprint doi: https://doi.org/10.1101/076554; this version posted September 21, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 4: LDA of item position in V1 using different subsets of voxels. Top row: all voxels from V1; middle row: only retinotopically activated voxels from V1; bottom row: only retinotopically suppressed voxels from V1.

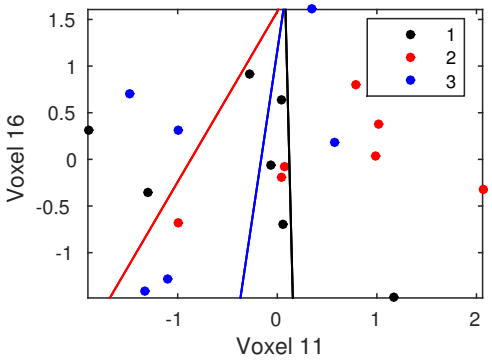
(A) V1: all voxels. LDA between-class boundaries based on the two most informative voxels.



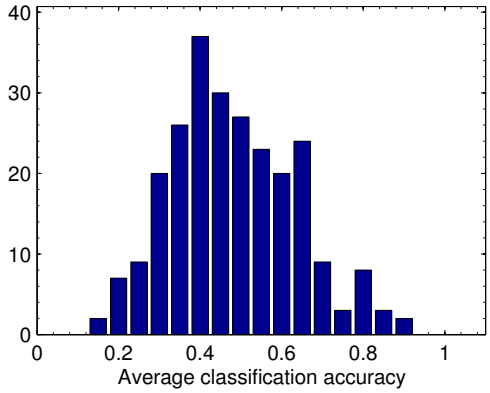
(B) V1: all voxels. Distribution of average LDA accuracy values (based on 250 simulations).



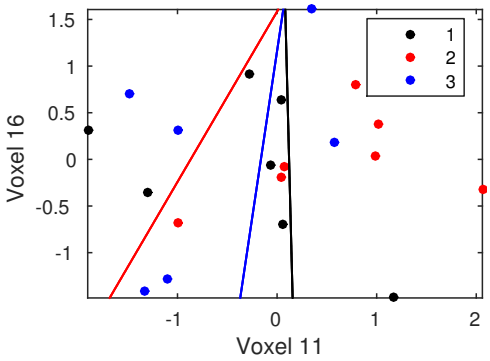
(C) V1: only activated voxels. LDA between-class boundaries based on the two most informative voxels.



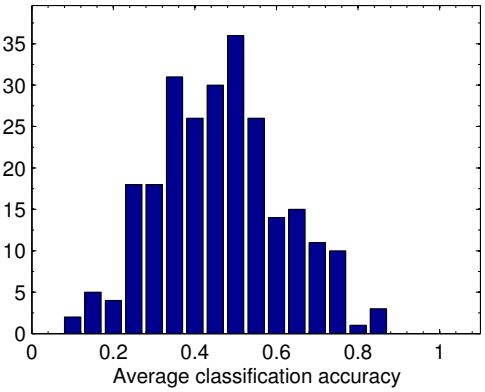
(D) V1: only activated voxels. Distribution of average LDA accuracy values (based on 250 simulations).



(E) V1: only suppressed voxels. LDA between-class boundaries based on the two most informative voxels.



(F) V1: only suppressed voxels. Distribution of average LDA accuracy values (based on 250 simulations).



bioRxiv preprint doi: https://doi.org/10.1101/076554; this version posted September 21, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Simulation of collinear processes

Here we simulate two types of position-collinear processes which can serve as a positional read-out. In the first case the brain area responds uniformly along the sequence (e.g. sensory adaptation) and in the second case units within the population respond differentially. We show that in the first case we can make reasonable *a priori* assumptions about the nature of the positional code and hence remove a uniform signal. However, when the population responds differentially to sequence positions there are no prior criteria to distinguish positional read-out from a positional code.

The MATLAB/Octave code for the simulated data and plots is freely available at http://imaging.mrc-cbu.cam.ac.uk/imaging/KristjanKalm/ovs/.

Uniformly changing signal across sequence positions

Here we model sensory adaptation in a simple sequence processing task as an example of a uniformly changing position-collinear process. We show how human fMRI data obtained with the same task fits the simulation results. We also propose a data pre-processing step – de-meaning of neural responses – as a tool to eliminate univariate signal collinear to the positional code.

Throughout the simulations we use the term 'brain region' for a population of neural units and the term 'voxels' for units themselves. This makes the terminology compatible with the experimental data presented from human fMRI experiments.

Representation of sequence items in a brain region

As a baseline condition we simulate the case where the only information stored in a brain region is item information without any positional code, and where there is no correlated information such as decay or interference. We simulate a sensory brain region of $n = 20$ voxels which encodes identities for three different items i as independent samples from the uniform distribution (Figure 5A):

bioRxiv preprint doi: https://doi.org/10.1101/076554; this version posted September 21, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$$\mathbf{r}_i \sim \mathcal{U}(0,1)$$

A brain region's response $\mathbf{Y}$ to the item i will be the item pattern $\mathbf{r}_i$ plus some noise sampled from n-dimensional Gaussian distribution with a zero mean.

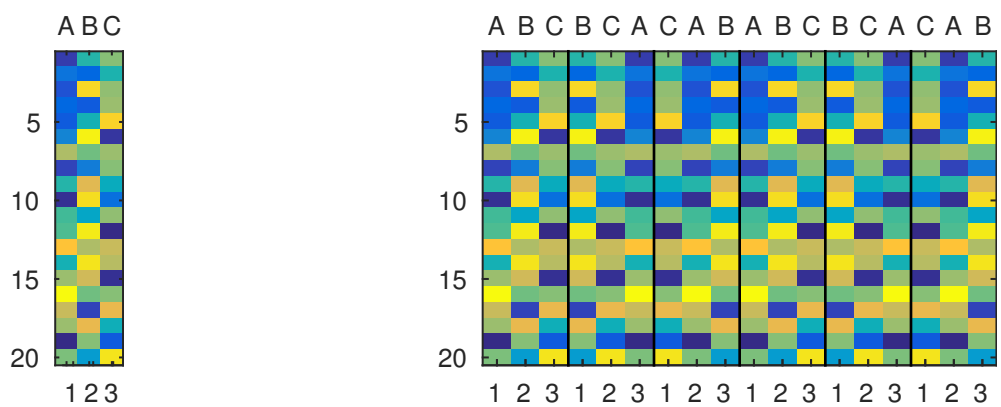
$$\mathbf{y}_i = \mathbf{r}_i + \mathcal{N}(0,\sigma)$$

To model a noisy average of these patterns we simulate an experiment where those three items are presented in different order as sequences for 6 times. The simulated response matrix $\mathbf{Y}$ depicts those 6 sequences with item and position values labelled on the x-axis (Figure 5B).

Figure 5: Simulated responses to items

(A) Item patterns over 30 voxels.

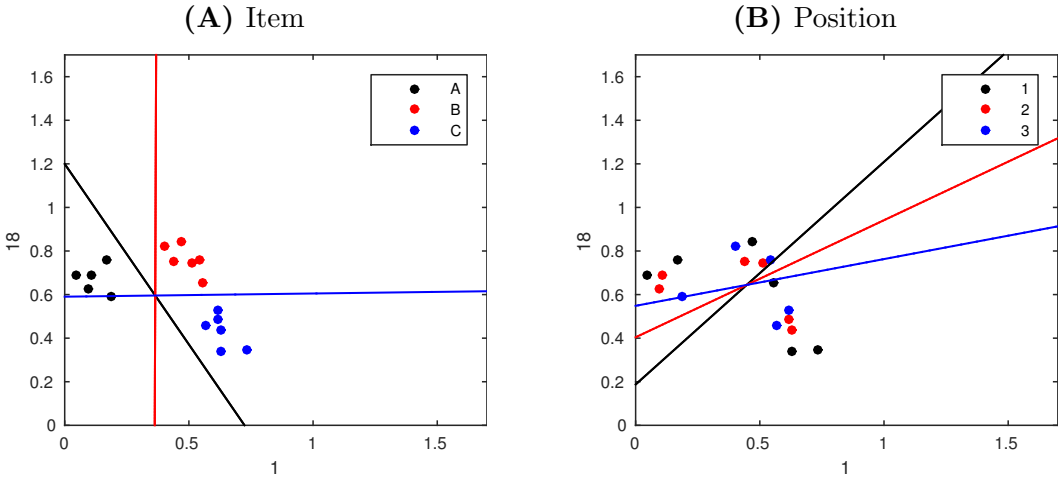
(B) Six sequences as permutations of three items. Item codes are displayed on the top of x-axis, position codes at the bottom.



As a result, the brain region's response matrix $\mathbf{Y}$ contains noisy representations of item identity but no information about position in the sequence. This can be visualised by plotting the scatter of the data $\mathbf{Y}$ and class borders (based on Fisher's Linear Discriminant Analysis, LDA) according to item and position labels (Figure 6). It is obvious that patterns $\mathbf{Y}$ are only linearly separable in terms of item identity (Figure 6A) but not position (Figure 6B).

bioRxiv preprint doi: https://doi.org/10.1101/076554; this version posted September 21, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 6: The scatter of item patterns and LDA between-class boundaries based on the two most informative voxels.



Sensory adaptation

So far we have assumed that item representations are completely independent of sequence position. Next we consider the case where there is a degree of sensory adaptation across the sequence. We simulate sensory adaptation for a brain region as a fixed vector across voxels multiplied by a decreasing function of sequence position, plus a Gaussian noise of fixed magnitude. This means that sensory adaptation will influence all voxels in the brain region similarly. In other words, in terms of a neural response of a brain region, sensory adaptation is a univariate signal decreasing monotonically over sequence positions.

We simulate sensory adaptation for all voxels i.e. voxels respond to stimulus positions $\{1, 2, 3\}$ by a decreasing vector $\mathbf{a} = [1, 0.7, 0.4]$. The average responses of the voxels can be shown as column-wise means of the response matrix (Figure 7B). As a result, the response of the brain region allows us to linearly separate both item identities and their positions in the sequence (Figure 7C). Re-running this simulation 250 times yields a distribution of average LDA accuracy values (Figure 7D).

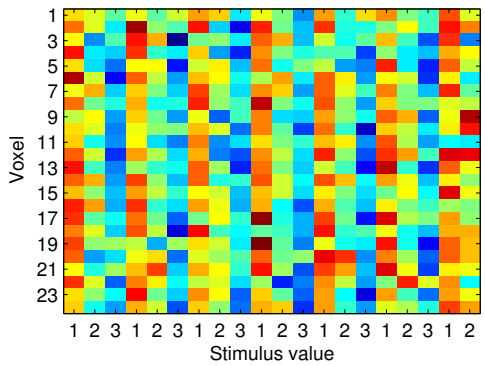
In sum, simulating sensory adaptation on top of independent item codes allows us to 'decode' items in terms of their positions since it is predicted by the amount of sensory adaptation. The same effect would result from any cognitive process collinear to the item position such as

bioRxiv preprint doi: https://doi.org/10.1101/076554; this version posted September 21, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

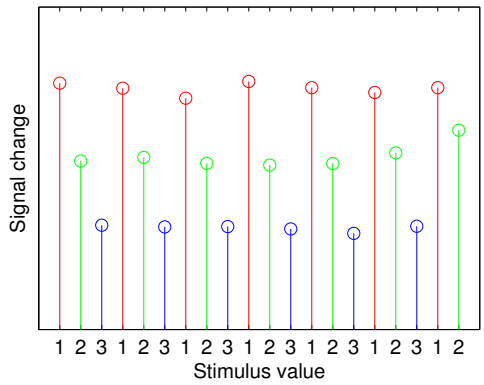
increasing memory load or passing time.

Figure 7: Simulation of sensory adaptation

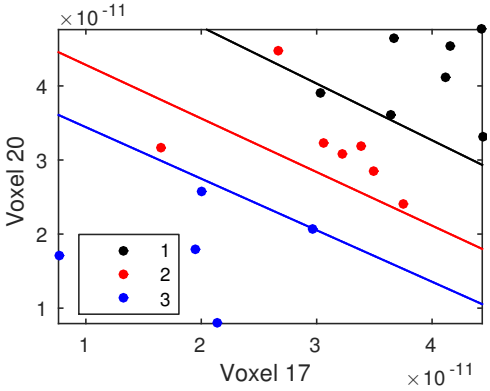
(A) Voxels' responses with sensory adaptation.



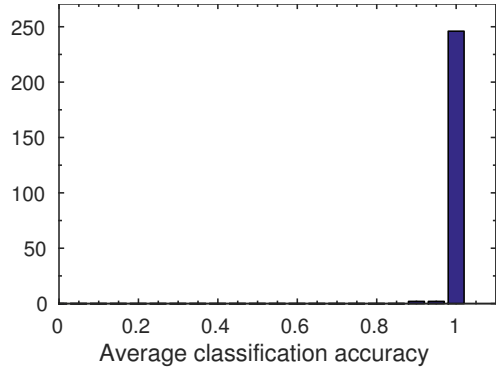
(B) Average responses of voxels as column-wise means of the response matrix.



(C) LDA between-class boundaries based on the two most informative voxels.



(D) Distribution of average LDA accuracy values (based on 250 simulations).



Eliminating uniform signal by de-meaning

If we assume that any collinear process to sequence position affects all voxels in the brain region uniformly then simple de-meaning of the response matrix will eliminate any univariate signal from the data.

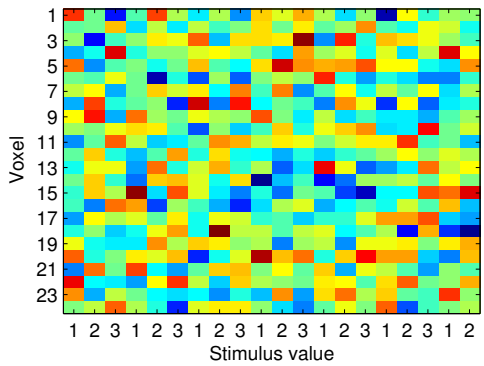
Here we z-score the response matrix before classification so that column-wise averages equal zero and values of the matrix correspond to z-scores based on the column mean. Carrying out LDA as before shows that the resulting average classification accuracy is at chance level as z-scoring the response matrix removes effects common to all voxels. Similarly, when z-scoring

bioRxiv preprint doi: https://doi.org/10.1101/076554; this version posted September 21, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

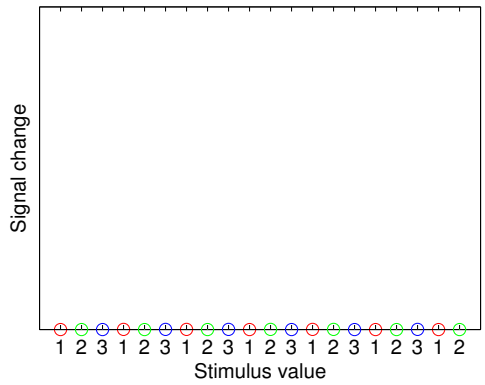
was applied to the fMRI data above, positional effects were no longer significant.

Figure 8: Simulation of sensory adaptation, z-scoring

(A) Voxels' responses with sensory adaptation.; z-scored.



(B) Average responses of voxels as column-wise means of the response matrix.

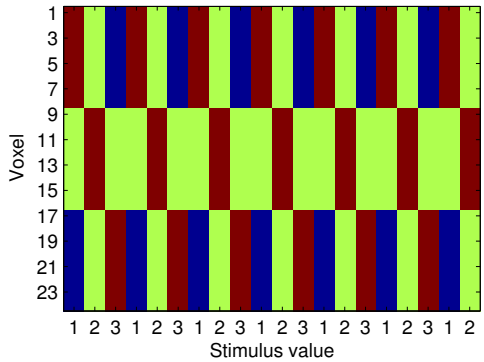


However, is is also possible that the sensory brain region actually contains a positional code. It is clear that in order to survive a de-meaning process a dedicated positional code must not be uniform across voxels. De-meaning process cannot affect a multivariate positional signal which affects voxels differentially. We can model each voxel's position preference T as a Gaussian likelihood function over the position values of the stimuli: i.e. each voxel responds most to a single position and less to adjacent positions (see the simulation code for different tuning distributions): $T \sim \mathcal{N}(Position, \sigma)$. Next, we add sensory adaptation, Gaussian noise, z-score the data, and carry out LDA, as above. The resulting average classification accuracy was close to 100%. Since z-scoring does not affect voxel pattern similarity, the positional code is used by the linear classifier to successfully distinguish between order positions.

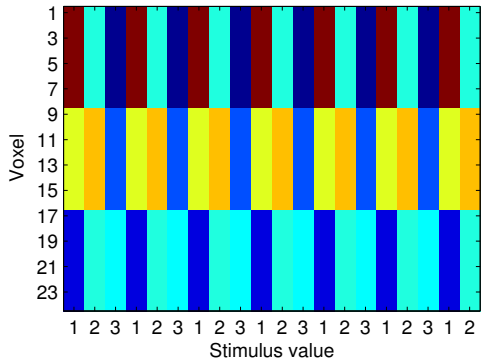
bioRxiv preprint doi: https://doi.org/10.1101/076554; this version posted September 21, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 9: Simulation of sensory adaptation

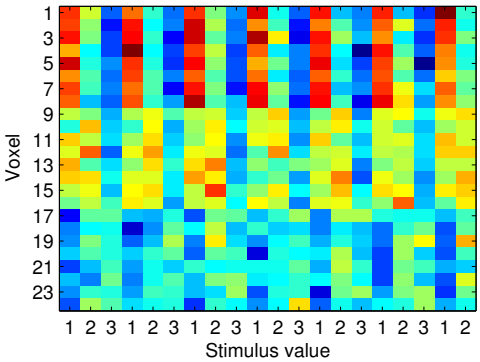
(A) Voxel response matrix based on positional preferences.



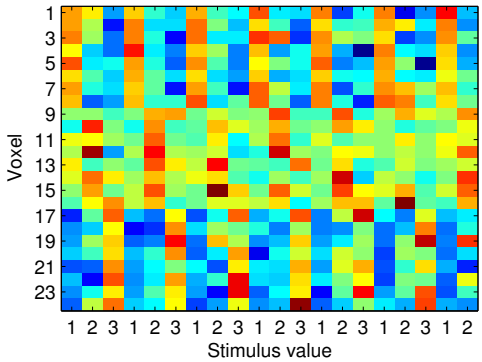
(B) Voxel response matrix: positional preferences + sensory adaptation.



(C) Voxel response matrix: positional preferences + sensory adaptation + Gaussian noise.



(D) Voxel response matrix z-scored. Column-wise means are zero.



However, any differential response within the brain region to sequence positions – such as the retinotopic activation example above – will similarly remain unaffected by z-scoring. As a result we can use de-meaning only to remove uniform effects from the brain region's response.

Summary of position-collinear effects

A number of cognitive processes take place while stimuli are processed in a sequence. Importantly, several of them – time, memory load, sensory adaptation – will be collinear to any signal tracking the position of items in a sequence.

We showed that uniform position-collinear processes – such as sensory adaptation– can be subtracted from neural responses by a de-meaning technique such as z-scoring. Importantly,

bioRxiv preprint doi: https://doi.org/10.1101/076554; this version posted September 21, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

this relies on an assumption that such processes will influence all units uniformly in a neural population. However, if individual voxels within a brain region respond differentially – such as in the case of retinotopic activation – the neural response becomes indistinguishable from a dedicated positional code.

Positional code from interference

A positional 'read-out' without a dedicated positional code can also arise from interference between sequentially presented stimulus representations. Here we use a simulation to show that a model of sequence representation which only includes item codes and no dedicated positional code can elicit positional effects given some interference between item codes. To illustrate this, imagine a brain region where the representations of successive items are overlayed on top of each other. Each successive item elicits a neural pattern that is a mixture of its own representation and a decaying representation of the preceding items. Such superimposed items could be linearly separable in terms of their positions alone without the need of any explicit representation of position.

Here we look at two cases of interference between item codes – additive and proportional interference– and how both can lead to position-like codes. The item representations are modelled exactly as above in Representation of sequence items in a brain region.

Additive interference

Interference between representations can occur when the state of the memory is not completely wiped clean every time a new stimulus arrives. Instead, the new state of the memory might be a mixture of the new stimulus and the previous state of the memory. Here we assume that at sequence position p the response of the brain region $\mathbf{Y}$ equals to the item pattern $\mathbf{r}_i$ plus some residual activity from the previous state of the brain region:

$$\mathbf{y}_p = \mathbf{r}_i + \pi \mathbf{y}_{p-1} \tag{1}$$

bioRxiv preprint doi: https://doi.org/10.1101/076554; this version posted September 21, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

where p is the position of the item in the sequence and π is the mixing coefficient which determines the proportion of the residual activity. Here π declines with a constant rate over previous states of $\mathbf{Y}$ so that:

$$\pi_p = \pi_0 \beta^{p-1} \tag{2}$$

where β is the rate parameter of the decreasing mixing coefficient π, and the initial value of $\pi_0 = 1$. Setting the initial value of π to 1 ensures that the current item pattern is always represented in full. To illustrate this mechanism consider two different β values and how they affect interference in a 3-item sequence 'CBA':

$$\begin{aligned}
\beta = 0.2 \quad & \mathbf{y}_{p=1} = \mathbf{r}_C \\
& \mathbf{y}_{p=2} = \mathbf{r}_B + 0.2\mathbf{r}_C \\
& \mathbf{y}_{p=3} = \mathbf{r}_A + 0.2\mathbf{r}_B + 0.04\mathbf{r}_C \\
\beta = 0.6 \quad & \mathbf{y}_{p=1} = \mathbf{r}_C \\
& \mathbf{y}_{p=2} = \mathbf{r}_B + 0.6\mathbf{r}_C \\
& \mathbf{y}_{p=3} = \mathbf{r}_A + 0.6\mathbf{r}_B + 0.36\mathbf{r}_C
\end{aligned}$$

It is clear that the size of the β parameter determines the amount of interference from previous items: when $\beta = 0$ there is no interference, and when $\beta > 1$ the activity from previous items contributes more to the current activity pattern $\mathbf{y}_p$ than the current item pattern $\mathbf{r}_i$.

Importantly, with each arriving item the overall activity of the brain region, as defined by the vector sum of $\mathbf{y}_p$, increases, since some of the previous response is added to the new response. In other words, additive interference as defined above (Eq. 1) guarantees that:

$$\sum \mathbf{y}_{p+1} > \sum \mathbf{y}_p$$

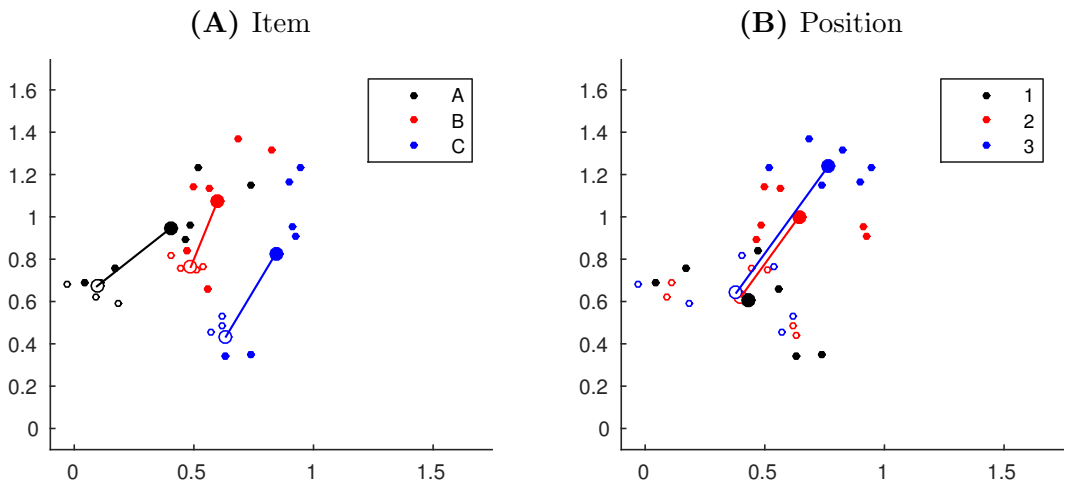
Similar increase in brain activity as a function of the number of sequentially presented items has been observed in several neuroimaging studies of short-term memory (Rottschy et al., 2012; Wager & Smith, 2003).

bioRxiv preprint doi: https://doi.org/10.1101/076554; this version posted September 21, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Additive interference enables position decoding

If we simulate additive interference as described above then despite the brain region only encoding item identity information we can linearly separate patterns $\mathbf{Y}$ in terms of their position because the total activity increases as a function of position. The effect of additive residual activity on sequence positions can be shown by plotting the positional means before and after interference transform (Figure 10).

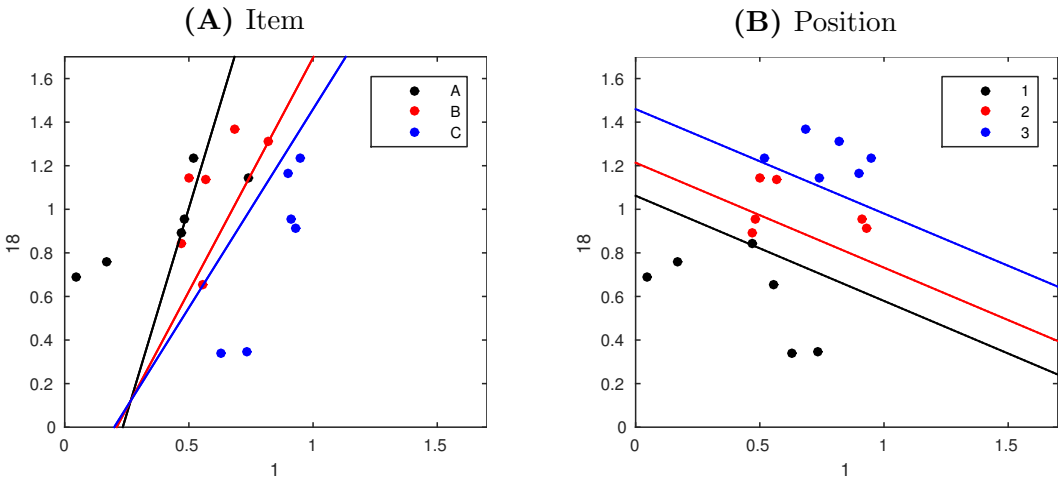
Figure 10: The transformation of response values for two voxels as a result of interference ($\beta = 0.5$). Small circular markers depict response patterns, larger circular markers depict pattern means. Empty markers depict the original patterns and means, filled markers depict the data after the interference. Solid lines depict the movement of the class means as a result of interference.



Note that patterns pertaining to the first positions in the sequence (black markers on Figure 10B) have not moved since there is no interference for the first items in the sequence from previous items. The position-wise transformation of the response patterns allows to separate them linearly using both item and position labels (Figure 11).

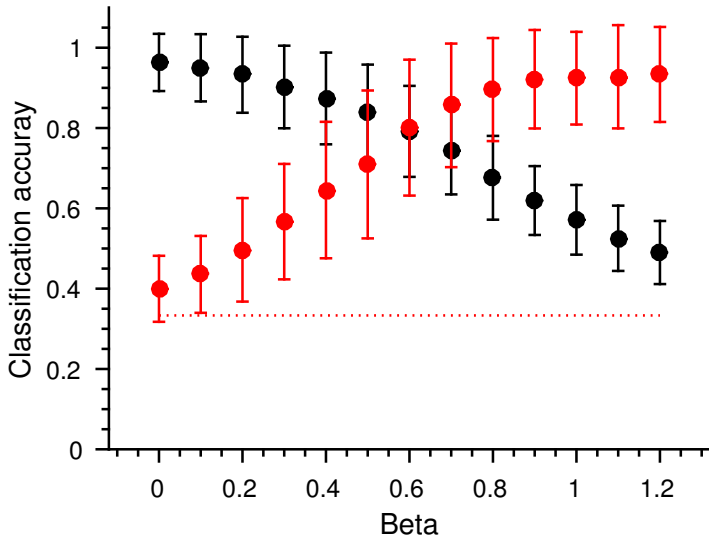
bioRxiv preprint doi: https://doi.org/10.1101/076554; this version posted September 21, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 11: LDA between-class boundaries for two voxels, interference $\beta = 0.5$



We can now decode the position of the items significantly above chance because item position correlates with the amount of response in the simulated brain region. Plotting the classification accuracy of both item and position as a function of interference (β parameter value) we can see that even with relatively small β values positional decoding becomes significantly greater than chance whilst it is always possible to decode item identity above chance (Figure 12).

Figure 12: Linear classification accuracy of item identity (black) and position (red) as a function of additive interference (β). The red dotted line shows chance level classification accuracy. Error bars depict SEM based on 1,000 simulations of the interference process with fixed parameter values.

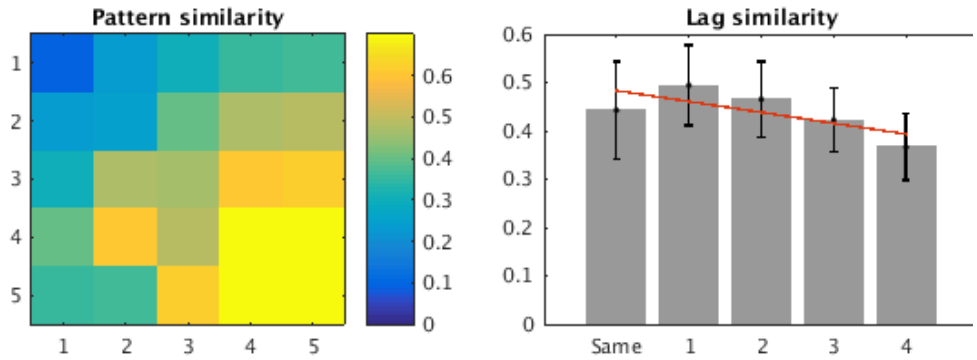


bioRxiv preprint doi: https://doi.org/10.1101/076554; this version posted September 21, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Positional pattern similarity decreases as a function of lag

Interference between item representations results in a change in pattern similarity across sequence positions. Specifically, between-position pattern similarity decreases as the distance between positions (lag) increases. In other words, pattern similarity is significantly higher across items that shared the same temporal position information than between items that are 1 or more positions apart (Figure 13). For the purposes of creating more positions the following plots display data generated exactly as above but with 5-item sequences instead of three.

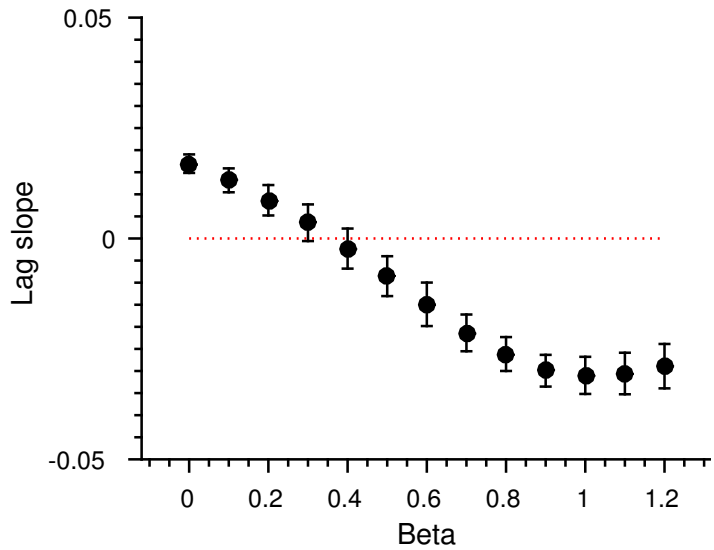
Figure 13: Similarity matrix on the left shows average positional pattern similarity, as measured by Pearson's ρ, based on additive interference with $\beta = 0.6$. Plot on the right visualises this similarity as a function of positional lag. The red line depicts a statistically significant negative slope over positional lag ($p < 0.005$).



Such an effect of positional pattern similarity has be observed in a number of animal and human studies (Devito & Eichenbaum, 2011; Fortin, Agster, & Eichenbaum, 2002; Hsieh, Gruber, Jenkins, & Ranganath, 2014; Hsieh & Ranganath, 2015) and interpreted as a signature of positional code. The size of the lag effect can be measured as the magnitude of the negative slope over lag values as depicted on Figure 13. Since positional effects are here solely caused by the interference mechanism it follows that the size of the lag effect correlates with the π parameter, which determines the extent of residual activity from the previous item. Figure 14 shows how the change in the lag effect as a function of additive interference.

bioRxiv preprint doi: https://doi.org/10.1101/076554; this version posted September 21, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 14: The size of the positional lag effect as a function of additive interference (β). Error bars depict SEM based on 1,000 simulations of the interference process with fixed parameter values.



Proportional interference

Pure additive interference is implausible since it presumes unlimited growth of the response in the brain region. We can cap the total response in the brain region ($\sum \mathbf{y}_p$) by normalising the response pattern every time a new item is presented. The easiest way to do this is to change the role β from the amount of residual activity to the *proportion* of residual activity. This requires a single change to the interference mechanism (Eq. 2) so that now we also weigh the current item representation $\mathbf{r}_i$, but with $1 - \pi$:

$$\mathbf{y}_p = (1 - \pi)\mathbf{r}_i + \pi \mathbf{y}_{p-1} \tag{3}$$

Whereas previously β represented the amount of interference from the previously presented item, now β determines the *proportion* of the previous item pattern $\mathbf{y}_{p-1}$ in the current item pattern $\mathbf{y}_p$. If we set $\beta = 0.2$, the representation of a four-item sequence A, B, C, D would evolve as follows:

$\mathbf{y}_{p=1} = \mathbf{r}_A$

$\mathbf{y}_{p=2} = 0.8\mathbf{r}_B + 0.2\mathbf{r}_A$

$\mathbf{y}_{p=3} = 0.8\mathbf{r}_C + 0.16\mathbf{r}_B + 0.04\mathbf{r}_A$

bioRxiv preprint doi: https://doi.org/10.1101/076554; this version posted September 21, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

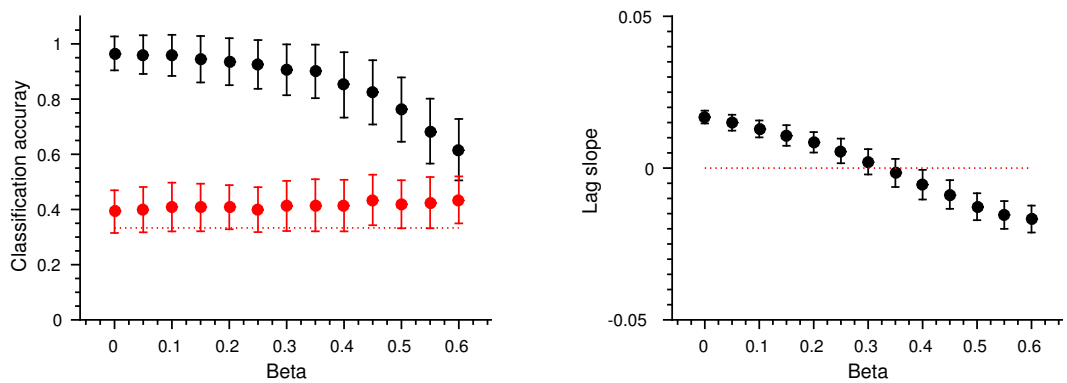
$$\mathbf{y}_{p=4} = 0.8\mathbf{r}_D + 0.16\mathbf{r}_C + 0.032\mathbf{r}_B + 0.008\mathbf{r}_A$$

Though the mechanism of interference here is the same as in the previous simulation, we no longer allow the response of the brain region ($\sum \mathbf{y}_p$) to grow as the sequence proceeds. In other words, we have eliminated any univariate signal correlated with position. Consequently, linear decoding of response patterns based on position is not significantly different from chance any more (Figure 15A, red line). However, the positional lag effect remains since it is based on pattern similarity (as measured by Pearson's ρ) which is insensitive to class means (Figure 15B).

Figure 15: Classification accuracy and positional similarity as simulated by the proportional interference mechanism. Error bars depict SEM based on 1,000 simulations of the interference process. Notice that β values on the x-axis have been approximately halved since the parameter now indicated the proportion of residual activity.

(A) Linear classification accuracy of item identity (black) and position (red) as a function of proportional interference (β). The red dotted line shows chance level classification accuracy.

(B) The size of the positional lag effect as a function of proportional interference (β).



In sum, de-meaning the neural response patterns only subtracts univariate effects of between-item interference. Pattern similarity effects of interference, such as the positional lag effect, still remain. It follows therefore that the positional lag effect alone is not a sufficient evidence for a neural positional code and additional statistical tests, such as classification analysis, are required.

bioRxiv preprint doi: https://doi.org/10.1101/076554; this version posted September 21, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Summary of item interference effects

Even in the absence of any true positional code, if the encoding of item information is based on overlaying item representations in a non-additive fashion this can potentially masquerade as a positional code. Depending on the magnitude of interference both position decoding and positional lag effects can be successfully simulated. Positional decoding is possible when residual activity from previous items is not capped and the brain region's mean response grows with sequence position. When the activity patterns are normalised so that the mean response stays the same then only the positional lag effect remains.

Other sources of interference

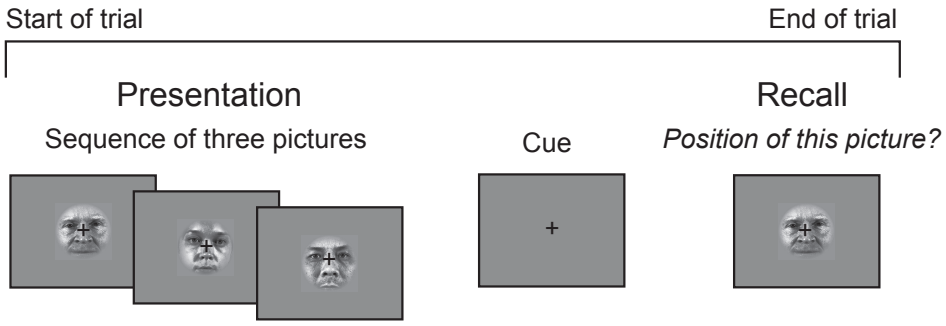
The mechanism of interference, as described above in the context of item codes, can be similarly applied to other variables of the experimental design. In fact, as outlined in the Introduction, any fixed parameter of the experimental design is collinear with positional effects. Next we briefly discuss how position-like codes emerge as a result of interference between task phases and as a result of temporally convolved measurement.

Interference between task phases

One of the most common tasks used in studying sequence representation is the *serial recall task* (Figure 16). In the serial recall task presentation of a sequence of items is usually followed by a response phase requiring the participant to recall the sequence. Importantly, the temporal order between task phases themselves is always fixed: recall must necessarily follow presentation, rest always occurs between the trials etc. As a result, the positional structure of the presented sequence in the task is collinear with the structure of the task itself. For example, in the serial recall task the last item in the sequence is always followed by the recall phase. Similarly, the first item in the sequence is always preceded by recall on the previous trial. As a result we can reliably predict the position of an item in the sequence based on its adjacency to different task phases.

bioRxiv preprint doi: https://doi.org/10.1101/076554; this version posted September 21, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

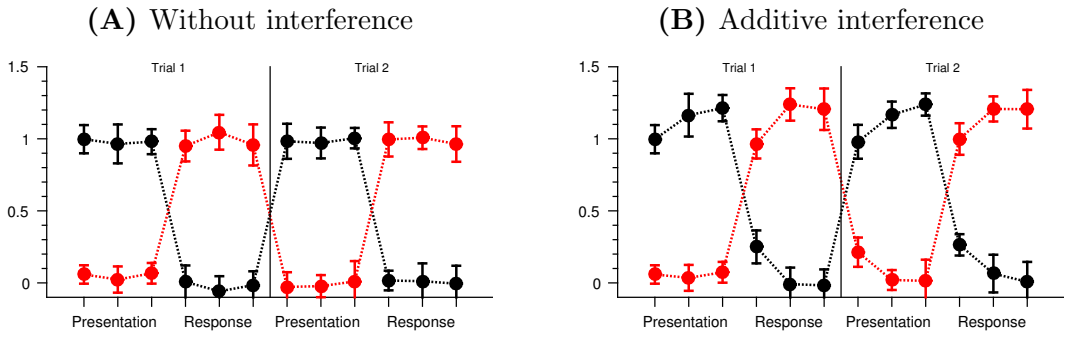
Figure 16: A serial recall task based on Kalm and Norris, 2016



We can model interference between task phases by simulating a response of 20 voxels as above, but during a single trial of a serial recall task. The task has two phases occurring in fixed order: *presentation* and *recall* (Figure 16). We assume that some voxels in the brain region are selective for the presentation and some for the recall phase. This selectivity can be described as voxels' likelihood to respond given a task phase. If there is no interference between task phases the response of phase-selective voxels is independent at any stage of the task: the previous phase of the task does not alter the voxels' activity at current stage (Figure 17A). However, if we implement Additive interference as described above then the extent of the response of phase-selective voxels becomes collinear with item position in the sequence (Figure 17B). Importantly, no item codes are necessary here, just sensitivity to task phases suffices. Due to interference we can now linearly separate the response patterns in terms of their sequence *position* because the total response changes as a function of task phase (Figure 17B). In every other aspect the mechanism is the same as described in Positional code from interference above.

bioRxiv preprint doi: https://doi.org/10.1101/076554; this version posted September 21, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 17: The average simulated activity of two sets of voxels, each sensitive either to the presentation or recall phase of the task. In this hypothetical task a presentation of three items in a sequence is followed by recall of three items.



Neurons' or voxels' sensitivity to a specific task phase is common, since in most experiments perceptual information is impossible to balance across task phases. For example, the presentation phase in serial recall task commonly uses a different stimulus modality (visual or auditory) than the following recall phase (manual or verbal recall). As a result, large patches of the cortex are only engaged during a specific phase of the task creating conditions described above.

Interference from measurement: functional MRI

So far we have described interference mechanisms arising between neural representations. However, equally importantly, interference between representations can result from noisy measurement. Similarly to representational interference, this can lead to positional effects which are spurious.

Functional MRI measures neural activity by detecting changes in the concentration of oxyhemoglobin and deoxyhemoglobin in neural tissue (BOLD signal). The relationship between a neural event and the corresponding BOLD signal can be described by a haemodynamic response function (HRF). Importantly, the HRF is non-linear and spread out over several seconds (Figure 18A), meaning that the BOLD signal corresponding to temporally adjacent events, such as items in a sequence or task phases, will always contain a response elicited by events preceding the event of interest (Figure 18B). This creates conditions similar to between-item and task

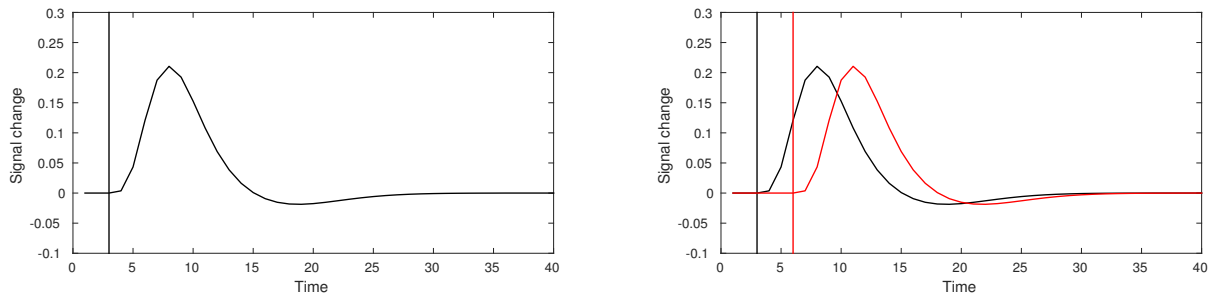
bioRxiv preprint doi: https://doi.org/10.1101/076554; this version posted September 21, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

phase interference described above – only this time there is no need for cognitive or representational interference. The temporal overlap in the BOLD signal will result in interference between measured item or phase representations even if the neural representations are independent of each other.

Figure 18: Temporal interference in fMRI

(A) The haemodynamic response function (HRF) with the vertical line representing the corresponding neural event.

(B) Temporal interference between two adjacent events: black and red lines.



Discussion

A major methodological barrier to studying the neural representation of positional code is that in any sequence processing task items in different position necessarily differ on other dimensions too. In this paper we used simulations and experimental data to show how both position-collinear experimental variables and interference between sequence items can lead to positional read-out in the absence of a dedicated positional code. This raises two important questions: (1) is it important to distinguish between a positional read-out and a 'true' positional code; and (2) what steps can be taken to delineate those in experimental data.

Positional read-out vs. dedicated positional code

Most models of neural sequence representation assume the existence of a dedicated positional code in the brain (Howard & Kahana, 2002; Henson & Burgess, 1997; Burgess & Hitch, 2006; Brown, Preece, & Hulme, 2000). However, since several cognitive processes (e.g. memory load, sensory adaptation) are collinear with any positional signal, a question arises whether those

bioRxiv preprint doi: https://doi.org/10.1101/076554; this version posted September 21, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

collinear processes could be used as a positional signal not just by the experimenter but also by the brain. We argue that a positional read-out from either simple position-collinear processes or between-item interference is not sufficient to support the storage and recall of a sequence.

Positional read-out from collinear processes is not sufficient for recalling a sequence

Recalling items in a sequence requires reinstating their order during recall. This problem is solved by positional models of sequence processing by associating each sequence item with its position during encoding and retrieving the order of items during recall by reinstating each positional code, which then cues the associated item (Figure 1B, e.g.: Burgess & Hitch, 2006; Howard & Kahana, 2002; Brown et al., 2000; Lee & Estes, 1981). However, it is hard to see how processes such as memory load or sensory adaptation could be used to cue associated items. Although experimenters can decode item position in a sequence based on memory load or sensory adaptation it is hard to see how "cortex as receiver" can use those neural signals to represent position and guide behaviour. For example, in case of using memory load (or any monotonically changing signal) as a positional code to cue associated items would require first to reinstate such 'load' to cue the corresponding item. However, such interpretation of 'memory load', which can be reinstated independently of the amount of items in memory, loses its original meaning and becomes a clumsy re-interpretation of a dedicated positional code. For this reason any *effect* of sequence processing, such as memory load or sensory adaptation, cannot be inverted into *cause* that would enable to associate items into sequence.

Positional read-out from between-item interference is not sufficient for recalling a sequence

We showed that interference between item representations can potentially masquerade as a positional code. This is because noisiness of the item representations changes monotonically over sequence positions as a result of interference. The change in the noise profile can therefore be used by the experimenter to reliably predict the position of the item in the sequence. However, as with simple position-collinear processes, it is hard to see how changes in the signal-to-noise

bioRxiv preprint doi: https://doi.org/10.1101/076554; this version posted September 21, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ratio can be used by the brain to guide sequence recall. The main consequence of interference via overlaying item representations is that the later items in the sequence are noisier than the early ones. This contradicts the well-established recency effect in sequence recall, where last items in the sequence are more accurately recalled than the ones in the middle (see Hurlstone, Hitch, & Baddeley, 2014, for a review). Secondly, using the noisiness of item representations as a positional code to cue associated items conflates the cause and effect relationship in sequence processing, as discussed above. The noisiness of the items would need to be reinstated independently of items themselves, thus necessitating the recoding of the noise levels into a dedicated positional signal.

Methods to dissociate between positional read-out and dedicated positional code

It is not possible to devise a task where the positional signal is orthogonal to other experimental variables since cognitive processes collinear to the positional code will always arise whenever stimuli are presented in sequence. However, the vast majority of previous studies on the positional code (Table 1) do not acknowledge the possibility of the 'contamination' of the positional code or take any measures to control for it.

Two assumptions are required to distinguish between a positional read-out and a 'true' positional code. First, position-collinear processes like memory load or sensory adaptation will uniformly affect all neural units engaged in encoding the item representations. This assumption is relatively uncontroversial if we presume that such processes are the result (and not source) of sequence processing. Second, we need to assume that a dedicated positional code is reflected in the position-sensitivity within a population of neural units. In other words, units respond selectively to sequence positions based on some tuning function. Under such conditions simple de-meaning (e.g z-scoring) of the neural data with respect to experimental condition (item position) will eliminate any univariate signal from the data including any univariate positional read-outs (see Eliminating uniform signal by de-meaning).

bioRxiv preprint doi: https://doi.org/10.1101/076554; this version posted September 21, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

However, we also showed that between-item interference can result in pattern similarity effects which masquerade as positional signal in the form of the lag effect (see Positional pattern similarity decreases as a function of lag and Proportional interference). Effects of pattern similarity are independent of signal amplitude and hence invariant to de-meaning. As a result, the effect of positional lag which has been used in several previous studies of positional code (Devito & Eichenbaum, 2011; Fortin et al., 2002; Hsieh et al., 2014; Hsieh & Ranganath, 2015) cannot be taken as a proof of neural positional code without ruling out between-item interference first. We show that this can be achieved by using linear classification analyses on the de-meaned neural responses.

Besides cognitive interference– such as based on overlaying item representations – positional read-out can result in noisy measurement, such as the temporal interference inherent in fMRI. In other words, any fMRI signal pertaining to successively presented sequence items will include a positional signal based on measurement error, even if we assume no interference between the neural representations of items themselves. As a result, the positional lag effect alone should never be used in fMRI studies as an indicator of neural positional representation (Hsieh et al., 2014; Hsieh & Ranganath, 2015).

Because the inherent temporal lag in the fMRI data sequentially presented stimuli will always be collinearly dependent on each other and suffer from confounds described above. Alternatively, a whole-sequence data can be used to extract positional information using the representational similarity analysis (Kriegeskorte, Mur, & Bandettini, 2008; Kalm & Norris, 2014).

Conclusions

In this paper we have explored two types of processes that could enable an experimenter to read out a positional 'code' in the absence of a dedicated positional code. First, we show that with any sequence processing task there are experimental variables collinear with the positional signal (e.g. time, memory load, etc.) which can serve as a positional code. Second, we show how interference between item representations, task phases, and measurement modalities can

bioRxiv preprint doi: https://doi.org/10.1101/076554; this version posted September 21, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

also lead to a similar positional read-outs.

We argue that it is important to distinguish between a positional read-out and a dedicated positional code, since only the latter has been shown to be compatible with experimental data. Furthermore, we argue that such collinear processes which enable positional read-out are the result of sequence representation not cause, and hence would not be able to even theoretically support sequence retrieval. Finally, we suggest practical steps in data analysis to distinguish between a positional read-out and a code. Furthermore, this paper shows that many results from behavioural and neural experiments studying the positional code must be treated with caution.

bioRxiv preprint doi: https://doi.org/10.1101/076554; this version posted September 21, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The MATLAB/Octave code for the simulated data and plots is freely available at http://imaging.mrc-cbu.cam.ac.uk/imaging/KristjanKalm/ovs/.

Table 1: Studies of neural representation of positional code

First author	Year	Stimuli	Task	Subject	Measurement
DuBrow	2014	visual	manual	human	fMRI
Averbeck	2003	motor	motor	monkey	electrophysiology
MacDonald	2011	odour	motor	rodent	electrophysiology
Averbeck	2007	visual	saccade	monkey	electrophysiology
Nieder	2012	visual	manual	monkey	electrophysiology
MacDonald	2013	odour	motor	rodent	electrophysiology
Mankin	2012	spatial	motor	rodent	electrophysiology
Manns	2007	odour	motor	rodent	electrophysiology
Rangel	2014	motor	motor	rodent	electrophysiology
Kraus	2013	motor	motor	rodent	electrophysiology
Nakajima	2009	manual	manual	monkey	electrophysiology
Berdyyeva	2011	visual	motor	monkey	electrophysiology
Hsieh	2014	visual	manual	human	fMRI
Berdyyeva	2010	visual	saccade	monkey	electrophysiology
DuBrow	2016	visual	manual	human	fMRI
Merchant	2013	auditory	manual	monkey	electrophysiology
Lehn	2009	visual	manual	human	fMRI
Hsieh	2015	visual	manual	human	fMRI
Isoda	2004	visual	saccade	monkey	electrophysiology
Averbeck	2006	visual	saccade	monkey	electrophysiology
Fujii	2005	visual	saccade	monkey	electrophysiology
Ginther	2011	odour	motor	rodent	electrophysiology
Hyde	2012	in vitro		rodent	electrophysiology
Inoue	2006	visual	manual	monkey	electrophysiology
Crowe	2014	manual	manual	monkey	electrophysiology
Naya	2011	visual	manual	monkey	electrophysiology
Manns	2007	odour	motor	rodent	electrophysiology
Gelfand	2003	auditory	manual	human	fMRI
Kalm	2014	auditory	auditory	human	fMRI
Ninokura	2004	visual	manual	monkey	electrophysiology
Barone	1989	visual	manual	monkey	electrophysiology
Amiez	2007	visual	manual	human	fMRI
German	2006	visual	manual	monkey	electrophysiology
Pastalkova	2008	motor	motor	rodent	electrophysiology

bioRxiv preprint doi: https://doi.org/10.1101/076554; this version posted September 21, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

Averbeck, B. B., Crowe, D. A., Chafee, M. V., & Georgopoulos, A. P. (2003). Neural activity in prefrontal cortex during copying geometrical shapes. II. Decoding shape segments from neural ensembles. *Experimental brain research*, *150*(2), 142–53. doi: 10.1007/s00221-003-1417-5

Berdyyeva, T., & Olson, C. (2010). Rank signals in four areas of macaque frontal cortex during selection of actions and objects in serial order. *Journal of neurophysiology*, *104*(1), 141–59. doi: 10.1152/jn.00639.2009

Berdyyeva, T., & Olson, C. (2011). Relation of ordinal position signals to the expectation of reward and passage of time in four areas of the macaque frontal cortex. *Journal of neurophysiology*, *105*(5), 2547–2559. doi: 10.1152/jn.00903.2010

Brown, G., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological review*, *107*(1), 127–81.

Burgess, N., & Hitch, G. (2006). A revised model of short-term memory and long-term learning of verbal sequences. *Journal of Memory and Language*, *55*(4), 627–652. doi: 10.1016/j.jml.2006.08.005

Devito, L., & Eichenbaum, H. (2011). Memory for the order of events in specific sequences: contributions of the hippocampus and medial prefrontal cortex. *The Journal of Neuroscience*, *31*(9), 3169–75. doi: 10.1523/JNEUROSCI.4202-10.2011

De-Wit, L., Alexander, D., Ekroll, V., & Wagemans, J. (2016). Is neuroimaging measuring information in the brain? *Psychonomic Bulletin & Review*. doi: 10.3758/s13423-016-1002-0

Fortin, N., Agster, K., & Eichenbaum, H. (2002). Critical role of the hippocampus in memory for sequences of events. *Nature Neuroscience*, *5*(5), 458–62. doi: 10.1038/nn834

Gouws, D., Alvarez, I., Watson, D. M., Uesaki, M., Rogers, J., & Morland, a. B. (2014). On the Role of Suppression in Spatial Attention: Evidence from Negative BOLD in Human Subcortical and Cortical Structures. *Journal of Neuroscience*, *34*(31), 10347–10360. doi:

bioRxiv preprint doi: https://doi.org/10.1101/076554; this version posted September 21, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

10.1523/JNEUROSCI.0164-14.2014

Henson, R., & Burgess, N. (1997). Representations of Serial Order. *The Quarterly Journal of Experimental Psychology*(789284387).

Henson, R., & Rugg, M. (2003, jan). Neural response suppression, haemodynamic repetition effects, and behavioural priming. *Neuropsychologia*, *41*(3), 263–270. doi: 10.1016/S0028-3932(02)00159-8

Howard, M., & Kahana, M. (2002). A Distributed Representation of Temporal Context. *Journal of Mathematical Psychology*, *46*, 269–299. doi: 10.1006/jmps.2001.1388

Hsieh, L.-T., Gruber, M. J., Jenkins, L. J., & Ranganath, C. (2014). Hippocampal activity patterns carry information about objects in temporal context. *Neuron*, *81*(5), 1165–78. doi: 10.1016/j.neuron.2014.01.015

Hsieh, L. T., & Ranganath, C. (2015). Cortical and subcortical contributions to sequence retrieval: Schematic coding of temporal context in the neocortical recollection network. *NeuroImage*, *121*, 78–90. doi: 10.1016/j.neuroimage.2015.07.040

Hurlstone, M. J., Hitch, G. J., & Baddeley, A. D. (2014). Memory for serial order across domains: An overview of the literature and directions for future research. *Psychological bulletin*, *140*(2), 339–73. doi: 10.1037/a0034221

Inoue, M., & Mikami, A. (2006). Prefrontal activity during serial probe reproduction task: encoding, mnemonic, and retrieval processes. *Journal of neurophysiology*, *95*(2), 1008–41. doi: 10.1152/jn.00552.2005

Kalm, K., & Norris, D. (2014). The representation of order information in auditory-verbal short-term memory. *The Journal of Neuroscience*, *34*, 6879–86. doi: 10.1523/JNEUROSCI.4104-13.2014

Kalm, K., & Norris, D. (2016). Representation of temporal order information in visual short term memory.

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in systems neuroscience*, *2*(November), 4. doi: 10.3389/neuro.06.004.2008

bioRxiv preprint doi: https://doi.org/10.1101/076554; this version posted September 21, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Larsson, J., & Smith, A. T. (2012). FMRI repetition suppression: Neuronal adaptation or stimulus expectation? *Cerebral Cortex*, *22*(3), 567–576. doi: 10.1093/cercor/bhr119

Lee, C., & Estes, W. (1981). Item and Order Information in Short-Term Memory: Evidence for Multilevel Perturbation Processes. *Journal of Experimental Psychology : Human Learning and Memory*, *7*(3), 149–169.

MacDonald, C., Carrow, S., Place, R., & Eichenbaum, H. (2013). Distinct Hippocampal Time Cell Sequences Represent Odor Memories in Immobilized Rats. *Journal of Neuroscience*, *33*(36), 14607–14616. doi: 10.1523/JNEUROSCI.1537-13.2013

MacDonald, C., Lepage, K., Eden, U., & Eichenbaum, H. (2011, aug). Hippocampal "Time Cells" Bridge the Gap in Memory for Discontiguous Events. *Neuron*, *71*(4), 737–749. doi: 10.1016/j.neuron.2011.07.012

Maier, A., Wilke, M., Aura, C., Zhu, C., Ye, F. Q., & Leopold, D. a. (2008). Divergence of fMRI and neural signals in V1 during perceptual suppression in the awake monkey. *Nature Neuroscience*, *11*(10), 1193–1200. doi: 10.1038/nn.2173

Mankin, E., Sparks, F., Slayyeh, B., Sutherland, R., Leutgeb, S., & Leutgeb, J. K. (2012). Neuronal code for extended time in the hippocampus. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(47), 19462–7. doi: 10.1073/pnas.1214107109

Manns, J. R., Howard, M. W., & Eichenbaum, H. (2007). Gradual changes in hippocampal activity support remembering the order of events. *Neuron*, *56*(3), 530–40. doi: 10.1016/j.neuron.2007.08.017

Nakajima, T., Hosaka, R., Mushiake, H., & Tanji, J. (2009). Covert representation of second-next movement in the pre-supplementary motor area of monkeys. *Journal of neurophysiology*, *101*(4), 1883–9. doi: 10.1152/jn.90636.2008

Naya, Y., & Suzuki, W. A. (2011). Integrating what and when across the primate medial temporal lobe. *Science*, *333*(6043), 773–6. doi: 10.1126/science.1206773

Page, M., & Norris, D. (1998). The Primacy Model: A New Model of Immediate Serial Recall. *Psychological Review*, *105*(4), 761–781.

bioRxiv preprint doi: https://doi.org/10.1101/076554; this version posted September 21, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Pastalkova, E., Itskov, V., Amarasingham, A., & Buzsáki, G. (2008, sep). Internally generated cell assembly sequences in the rat hippocampus. *Science*, *321*(5894), 1322–7. doi: 10.1126/science.1159775

Rottschy, C., Langner, R., Dogan, I., Reetz, K., Laird, A. R., Schulz, J. B., ... Eickhoff, S. B. (2012). Modelling neural correlates of working memory: A coordinate-based meta-analysis. *NeuroImage*, *60*(1), 830–846. doi: 10.1016/j.neuroimage.2011.11.050

Smith, T., Singh, K., & Greenlee, M. (2000). Attentional suppression of activity in the human visual cortex. *Neuroreport*, *11*(2), 271–277. doi: 10.1097/00001756-200002070-00010

Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M.-M., & Egner, T. (2008, sep). Neural repetition suppression reflects fulfilled perceptual expectations. *Nature Neuroscience*, *11*(9), 1004–1006. doi: 10.1038/nn.2163

Wager, T., & Smith, E. (2003). Neuroimaging studies of working memory: A meta-analysis. *Behavioral Neuroscience*, *3*(4), 255–274.