A Cognitively Plausible Model of Visual Working Memory

Anna Penzkofer (anna.penzkofer@vis.uni-stuttgart.de)
University of Stuttgart, Institute for Visualization and Interactive Systems (VIS), Germany

P. Michael Furlong (michael.furlong@uwaterloo.ca)
University of Waterloo, Centre for Theoretical Neuroscience (CTN), Canada

Chris Eliasmith (celiasmith@uwaterloo.ca)
University of Waterloo, Centre for Theoretical Neuroscience (CTN), Canada

Andreas Bulling (andreas.bulling@vis.uni-stuttgart.de)
University of Stuttgart, Institute for Visualization and Interactive Systems (VIS), Germany

Abstract

Visual working memory (VWM) plays a fundamental role in cognitive processes, such as perception, attention, and reasoning. However, existing approaches to modelling VWM are not integrated into cognitive architectures and lack interpretability with respect to their parameters. To address this limitation, we propose a novel VWM model based on the well-established Semantic Pointer Architecture (SPA). In contrast to previous works, our model is the first to integrate a VWM model with a cognitive attention model. It only requires three interpretable hyper-parameters: spatial capacity, feature certainty, and memory decay. We experimentally show that our base model without memory decay replicates the set-size effect and swap errors of human data on a continuous reproduction task. More importantly, we show that by introducing a memory decay, we can achieve a statistically significant ($p \ll 0.001$) improvement in model fit, suggesting a potentially important role of memory decay in VWM. Further, our VWM model can be easily extended to model pre- and post-cue conditions, consistently achieving KL divergence between modelled and human performance of less than 0.05.

Keywords: visual working memory, vector symbolic algebra, spatial semantic pointer, colour reproduction task

Introduction

Working memory functions as a cognitive workspace, maintaining representations essential for reasoning about temporally relevant stimuli, context, and actions (Chai, Abd Hamid, & Abdullah, 2018). *Visual* working memory (VWM) supports the processing of visual information and requires a computational model capable of accounting for observed behavioural phenomena. To enable future integration into cognitive models, a VWM model should be embedded within a comprehensive cognitive modelling framework. Ultimately, such a model should also be translatable into spiking neural networks to explain human neural data.

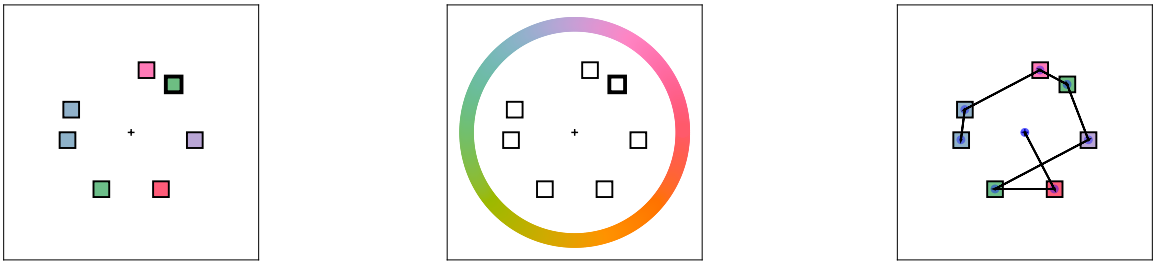
In this paper, we present a VWM model based on the well-established Semantic Pointer Architecture (SPA; Eliasmith, 2013), effectively integrating a VWM model into a cognitive architecture. The SPA employs Spatial Semantic Pointers (SSPs; Komer, Stewart, Voelker, & Eliasmith, 2019; Dumont & Eliasmith, 2020), which are embeddings of continuous variables that together with algebraic statements from Plate's Holographic Reduced Representations (HRR; Plate, 2003) form a closed algebra. This framework enables a full translation into spiking neural networks in the future (*e.g.* Eliasmith et al. (2012); Komer et al. (2019)). We evaluate our VWM model on a continuous colour reproduction task with experimental data collected by Oberauer & Lin (2017) and achieve a good model fit for different variants of our model. Our model is similar to previous SPA-based working memory models (Choo & Eliasmith, 2010; Gosmann & Eliasmith, 2015) but indexes memory contents using continuous spatial locations instead of discrete integer slots.

Early Models of VWM. Early methods modelled VWM as a fixed number of slots, with a capacity of about four objects (*e.g.*, Luck & Vogel, 1997). Luck & Vogel (1997) also found that these objects stored conjunctions of features, or *bound* representations. Alvarez & Cavanagh (2004) proposed that the allocation of fixed resources best described VWM, and that information-dense visual stimuli overload the capacity of VWM, a view supported by Bays & Husain (2008). Zhang & Luck (2008) fused these two methods into a "Slot-Averaging" model, improving the results. The Variable-Precision (VP) model of van den Berg, Shin, Chou, George, & Ma (2012), proposed that working memory allocates resources in a way that varies around an average and that is itself a function of the number of objects in memory. The VP model includes the limited resource model by Bays & Husain (2008) as a special case.

Distributional Models. Instead of modelling objects as occupying slots, Signal Detection Theory (SDT; Verghese, 2001) posits each stimulus is encoded by a corresponding noisy channel – cortical representations of continuous embeddings of stimuli. Wilken & Ma (2004) proposed that it is the presence of noisy signals and not the absence of resources that leads (probabilistically) to the misidentification of targets and replicated capacity observations in VWM. This method is compatible with our approach, where our distributed embeddings of stimuli can be understood probabilistically (Furlong & Eliasmith, 2023). To address the question of storing conjunctions of features, Schneegans & Bays (2017) presented a model of VWM that binds location with visual characteristics of presented stimuli. Their method of implementing conjunctive coding is reminiscent of the Dynamic Neural Fields theory (Schöner & Spencer, 2016) approach to neural population encoding. Likewise, our binding operation between distributed representations can be implemented neurally (Komer et al., 2019), i.e., it can support conjunctive coding as well.

Interference Model. Oberauer & Lin (2017) proposed the interference model (IM), which explicitly defines a bivariate feature and context distribution, creating a two-dimensional

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*



(a) (b) (c)

Figure 1: Colour reproduction task – the stimulus consists of differently coloured squares **(a)**, where the highlighted location in the test screen **(b)** is probed for selection on the colour wheel; **(c)** shows an example scan path as predicted by EMMA.

binding space. Oberauer & Lin (2017) provided human responses on the continuous colour reproduction task, introduced by Wilken & Ma (2004), with additional experiments including pre- and post-cues. The IM was able to capture the set size effect, where human error increases with increasing set size. The IM also fitted the distribution of non-target responses, which indicates the occurrence of swap errors in VWM, while both the Slot-Averaging (Zhang & Luck, 2008) and VP model (Bays & Husain, 2008) did not.

Our proposed VWM model is similar to the IM and other models, as it also binds the context domain and feature domain of objects. However, we do not define specific distributions, and our model requires fewer hyper-parameters, which are also more interpretable. In the following, we present our model with additional variants and evaluate them on the three experiments used by Oberauer & Lin (2017). We find that our model captures the set size effect, as well as swap errors. Most interestingly, we find that a memory decay parameter significantly increases model fit to human performance, indicating its importance in modelling VWM.

Method

We constructed our models using a Vector Symbolic Algebra (VSA; Gayler, 2004). VSAs are dimensionality-preserving algebras over high-dimensional vectors that represent structured data in a distributed manner. The VSA we use here comes with four main operators: *similarity* ($\phi \cdot \psi$), which is used to compare two vectors; *bundling* ($\phi + \psi$), which is used to create the superposition of two vectors; *binding* ($\phi \otimes \psi$), which creates a new vector that represents a conjunction of two vectors; *inverse* (ϕ^{-1}), which creates a new vector that is an approximate multiplicative inverse, *i.e.*, $\phi \otimes \psi \otimes \phi^{-1} \approx \psi$. Following work on Spatial Semantic Pointers (SSP; Komer et al., 2019), we used this VSA to embed continuous-valued data by defining:

$$\phi_\lambda(\mathbf{x}) = \mathcal{F}^{-1}\left\{e^{iA\lambda^{-1}\mathbf{x}}\right\}.$$

where $\mathbf{x} \in \mathbb{R}^m$ is the embedded data, $A \in \mathbb{R}^{d\times m}$ is the *phase matrix*, λ is the length scale parameter, the exponentiation is applied element-wise to $A\lambda^{-1}\mathbf{x}$, and $\mathcal{F}^{-1}$ is the inverse Fourier transform. We used Plate's Holographic Reduced Representations (HRR) (Plate, 2003) that uses circular convolution $\circledast$ for binding, vector addition for bundling, and the vector dot product for similarity. Previous work showed that these vector embeddings of data admit a probabilistic interpretation (Furlong & Eliasmith, 2023), which can support uncertainty representations found in other working memory models (Bays, Schneegans, Ma, & Brady, 2024).

To build a representation of a given visual stimulus, we encoded all stimuli in the scene sequentially. We encoded a single object by binding its feature (colour) $\phi_{\lambda_c}(c)$ to its spatial location $\phi_{\lambda_g}(x,y)$, which can also represent a vector of multiple locations. The full scene was then encoded as:

$$M = \sum_i \delta(i) \cdot \left[\phi_{\lambda_c}(c_i) \circledast \phi_{\lambda_g}(x_i, y_i)\right], \tag{1}$$

where δ is an adjustable encoding factor based on the decay parameter γ. This approach can be interpreted as using a slot-filling representation with continuous slots. The model's capacity is controlled by the degree of smoothing between continuous values, i.e., scale factors on the arguments c_i, x_i, and y_i that affect the respective feature and spatial certainty.

Feature Certainty. In our model, the feature certainty is modelled by the length scale parameter λ_c. While we encode colour on a colour circle, i.e., in 360 degrees, $\lambda_c \in [0.5 \cdot \frac{180}{\pi}, 2.5 \cdot \frac{180}{\pi}]$ effectively scales colour to radians and achieves smooth transitions between neighbouring colours. Increasing λ_c corresponds to a higher uncertainty in the feature dimension (see Figure 4). In other words, colours are more easily misremembered for adjacent colours.

Spatial Certainty. Similarly, the spatial certainty is modelled by the length scale parameter λ_g for the spatial dimensions x and y. When encoding locations λ_g affects the encoding resolution, i.e., a higher λ_g decreases resolution in spatial dimensions, letting neighbouring points be remembered as one (see Figure 4). Hence, λ_g models spatial uncertainty and allows for replicating swap errors between close objects.

Memory Decay. We set the memory decay as $\delta(i) = \gamma^i$, with γ as an adjustable parameter. In the base model **SSP-base**, the decay parameter was $\gamma = 1$, effectively disabling

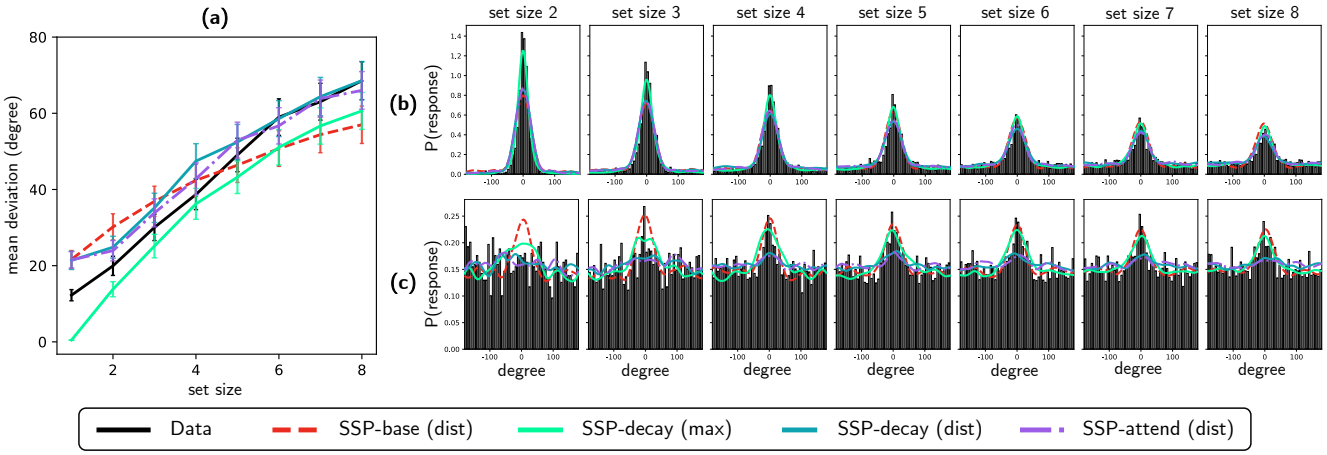


Figure 2: Results of base model: showcasing set-size effect (a), distribution fit of target responses (b), and distribution fit of centred non-target responses (c).

the memory decay. For the **SSP-decay** model, we set $\gamma < 1$ and assign each object a random position i. The last seen object was at $i = 0$ and, therefore, received $\gamma^i = 1$ as an encoding factor, while previously seen objects $i = 1, \ldots, n$ were multiplied by an exponentially smaller factor δ.

Attention-based Model

To estimate human scan paths, we integrated the cognitive attention module EMMA (Salvucci, 2001) – among the most popular modules for the ACT-R cognitive architecture (Anderson, Matessa, & Lebiere, 1997). EMMA takes bounding box locations of objects as input to predict a human-like scan path. Scan paths consist of consecutive saccades with a time of encoding the respective objects in between. The object encoding time depends on their frequency f_i and the eccentricity ε_i, i.e., the distance between the last eye position to the new object in visual angle, computed as $T_{\text{enc}} = K \cdot -\log f_i \cdot e^{k\varepsilon_i}$. K is a constant eye movement scale factor, while k is the eye movement angle parameter. For both, we used the default parameters in pyactr[1].We enabled the centre bias in the EMMA module, as a centre bias is encouraged in human trials by displaying a central fixation cross (Oberauer & Lin, 2017).

With the EMMA module, we predicted a unique scan path for each trial and integrated it into our SSP model (**SSP-attend**). An example scan path prediction can be seen in Figure 1 **(c)**. With the predicted order of encoded objects and their respective timestamps t_i, we computed the memory decay $\delta(i) = \gamma^{t_i / t_{\text{avg}}}$ and divided by the maximum δ, so that the last seen object has no decay and previous objects had a similar exponential decay to the decay model. Hence, in our experiments, the key difference between SSP-decay and SSP-attend was the predicted scan paths. For experiments with more complex objects that require a varying amount of encoding time T_{enc}, the memory decay calculation would differ more.

[1] https://github.com/jakdot/pyactr

Memory Decoding

To probe the spatial semantic pointer (SSP) memory M at the target location (x, y), we used unbinding, i.e., binding with the approximate inverse $M \circledast \phi(x, y)^{-1} = \psi$. This resulted in a new SSP ψ that we compared to a vocabulary of all colours encoded as SSPs with the similarity operator, yielding a similarity distribution across all colours. We converted similarity to a probability distribution over colour, using Born's rule, and then either picked the colour with the maximum similarity to ψ as our response, i.e., the maximum likelihood predictor, or we drew a random sample from the distribution. Drawing a sample from the distribution introduces a prediction error, automatically geared towards predicting non-target colours. In the following, we present the best results achieved on both response types for all models.

Hyperparameters

For all models, we performed a hyperparameter search to find the best combination of length scales λ_g, λ_c, and memory decay γ, evaluated with a coarse grid search for the overall model fit to the human data of Experiment 1. We selected λ_g to be between 10 and 500, λ_c between 30 and 150, which corresponds to 0.5 to 2.5 in radians, and γ from 0.4 to 1.0. The best parameters for each model, evaluated using KL divergence to the human response distribution, are summarised in Table 1. We optimised both the base model, where $\gamma = 1$ is fixed to disable the memory decay, and the SSP-decay model with variable γ, for both response types. The δ_{cue} parameter is only necessary for experiment two, where we changed the decay factor for the object in the pre-cue location.

Pre- & Post-cue Conditions

Oberauer & Lin (2017) evaluated the effect of pre- and post-cues in the continuous colour reproduction task. They found that the average error decreased for the pre-cue condition and barely changed for the post-cue condition. We extended our model to account for pre-cues by adjusting the factor δ

Model	Response	λ_g	λ_c	γ	δ_{cue}
SSP-base	max	390	110	1	2
SSP-base	dist	270	60	1	2
SSP-decay	max	230	150	0.6	0.67
SSP-decay	dist	90	60	0.4	0.67

Table 1: Best hyper-parameters for base and decay models with maximum and distribution response type.

that multiplies the bound colour and location SSPs. More specifically, we increased the factor for the cued location to $\delta(i_{\text{cue}}) = 2$ for the base models, which otherwise had a factor of $\delta(i) = 1$ for all objects i. This effectively added the cued location into memory twice and, therefore, increased its likelihood of being recalled. For the SSP-decay models, we set $\delta(i_{\text{cue}}) = 0.67$ to match the probability of receiving a valid cue, where only the last seen object was encoded with $\delta = 1$ and all other objects exhibited a strong exponential decay. For consistency and to highlight the generalisation ability of our model, we kept all other hyperparameters fixed.

Experimental Design

To directly compare our model with the IM proposed by Oberauer & Lin (2017), we evaluate our model on the continuous colour reproduction task as originally described in Wilken & Ma (2004). In the colour reproduction task, participants were asked to memorise up to eight coloured squares (see Figure 1 (a)). After a blank screen was shown for one second, they saw the probing screen in Figure 1 (b), where the target location was marked with a thicker border. The continuous colour wheel was displayed, and participants were asked to select the colour they remembered from this location.

Oberauer & Lin (2017) collected data on the colour reproduction task and observed interesting phenomena: (1) the set size effect: reproduction error increases with an increasing number of items, (2) the distribution of non-target items, i.e., it is more likely that participants remember a colour of the non-target items instead of a randomly selected one, which can also be defined as a function of the distance in both feature and spatial dimension, and (3) a focus of attention, where a pre-cue of the probed location significantly increases reproduction accuracy.

We used the implementation by Oberauer & Lin (2017) as they provide human data from 20-21 participants for three different experiments: In **Experiment 1**, the set size of the stimulus was varied from 1-8 with 100 trials per size. In **Experiment 2**, the colours were slightly altered, and a pre-cue was presented before the stimuli were shown. In 67% of trials, the pre-cue was valid and indicated the location that was later probed. Participants were instructed on the possibility of false pre-cues. **Experiment 3** was the same as Experiment 2, but with post-cues, i.e., the cue appeared after the one-second blank screen, right before the probing screen.

Results

Figure 2 summarises the results of Experiment 1. All models replicate the set size effect found in human data (see Figure 2 (a)). Comparing this to Figure 3 in Oberauer & Lin (2017), we find our model fits the set size effect on par with their Interference Model (IM). All distribution response models slightly over-predict the error until set size five. In contrast, the maximum response model – SSP-decay (max) – generally under-predicts the error, with almost zero error when the set size is one, making it more aligned with the human data until set size five. At this point, the distribution response model fits better. The best fit of the set size effect can be observed with the SSP-attend model, which incorporates the attention model EMMA (Salvucci, 2001). Interestingly, all distribution response models struggle with correctly predicting the error for a set size of one. Here, human error may arise from difficulties in distinguishing or selecting among 360 colours. We argue that these motor and perceptual factors are hard to incorporate into our visual memory model.

Following Oberauer & Lin's analysis, we computed a centred target response, shifting human and model responses so that the correct answer was always displayed at zero degrees. In Figure 2 (b), a histogram of the human responses for Experiment 1 is shown in black, and our models' density functions, estimated as a von Mises distribution ($\kappa = 20$), are overlayed. Similarly, we present the non-target responses in Figure 2 (c). Here, we compute the centred responses for all non-targets (up to seven) and find that objects with colours of non-targets are often recalled erroneously, starting at a set size of four. The SSP-base (dist) model accurately replicates the human response distribution for set sizes starting at five. However, for lower set sizes, it over-predicts the non-target distribution and, therefore, under-predicts the target distribution. The maximum response decay model – SSP-decay (max) – best fits the target response distribution but significantly over-predicts the non-target response. This is likely due to the high uncertainty in both the spatial dimension governed by λ_g and the feature dimension governed by λ_c, which were fine-tuned to yield the best possible average KL divergence for this model. The SSP-decay (dist) model still captures the non-target response effect (Figure 2 (c)) but is more moderate and correctly predicts a more uniform distribution for set sizes two and three. The attention-based model SSP-attend (dist) is very similar to the SSP-decay (dist) model; however, it has a slightly lower non-target prediction.

Overall, the distribution response models are better regarding KL divergence for both the target and non-target distributions. We compare results in Table 2 with KL divergence of all models' target response distribution and non-target response distribution compared to human data. To minimise random effects, we take the average across five random seeds `[0, 1, 2, 3, 4]` and report the standard deviation in brackets. We find that the SSP-decay model performs significantly better than the base model for both the maximum response as well as the distribution response. In fact, an inde-

		Experiment 1			Experiment 2	Experiment 3
Model	**Response**	**Target**	**Non-target**	**Average**	**Average**	**Average**
SSP-base	max	0.083 (0.004)	0.048 (0.002)	0.066 (0.002)	0.041 (0.001)	0.044 (0.001)
SSP-base	dist	0.064 (0.001)	0.015 (0.001)	0.040 (0.001)	0.035 (0.001)	0.032 (0.001)
SSP-decay	max	0.064 (0.004)	0.013 (0.001)	0.039 (0.002)	<u>0.030</u> (0.003)	0.024 (0.003)
SSP-decay	dist	<u>0.045</u> (0.002)	<u>0.012</u> (0.001)	<u>0.028</u> (0.001)	0.031 (0.002)	**0.021** (0.001)
SSP-attend	max	0.062 (0.007)	0.013 (0.000)	0.037 (0.004)	**0.030** (0.003)	0.024 (0.003)
SSP-attend	dist	**0.044** (0.002)	**0.012** (0.001)	**0.028** (0.001)	0.031 (0.002)	<u>0.022</u> (0.001)

Table 2: KL divergence mean (std) for models' predictions to target and non-target distribution across five random seeds. Bold numbers indicate the best values, while underlined numbers indicate the second-best.

pendent two-tailed t-test shows a statistically significant difference ($p \ll 0.001$) in both cases. However, the additional improvement by SSP-attend over the SSP-decay model is not significant in either case ($p \approx 0.35$), as particularly for the maximum response, we see an increase in standard deviation.

Experiments 2 & 3

In Experiment 2, where pre-cues were presented to participants before the stimuli, the overall response mean deviation decreased compared to Experiment 1. Consequently, the set size effect curve decreased in slope. With our extended SSP models, we observed the same effect as in Experiment 1, where the maximum response models slightly under-predicted the mean deviation and the distribution response models over-predicted up until set size six, where they fit perfectly. Interestingly, the maximum response variants of SSP-decay and SSP-attend were slightly better than the distribution versions in this experiment (see Table 2).

In Experiment 3, post-cues were used but were found not to have a significant effect on human response distributions. Both the set size effect and the swap errors with non-target responses are observed in the data. Our models capture the effects without any fine-tuning of hyper-parameters; in fact, they achieve even lower KL divergence in Experiment 3 than in the others. This might be due to the fact that in Experiment 2 and 3, only set sizes of 1, 2, 4, 6, and 8 were presented.

Parameter Sensitivity & Interpretation

We evaluated the parameter sensitivity of our three model parameters on Experiment 1. In Figure 3, we plot the variability of KL divergence of all parameters compared to the variability introduced by random sampling, i.e. the random seeds. We find that the memory decay γ has the most variability. The length scale parameter λ_c, which affects feature certainty, also exhibits some variability, followed by λ_g, the grid length scale with the lowest variability. However, all parameters have a larger effect than the different random seeds.

We further analysed parameter sensitivity by running a linear regression model predicting the average KL divergence score based on the given parameters. Overall, the regression model achieved an R^2 score of 0.74, i.e. explains 74% of the variance. The individual parameters, grid length scale λ_g, colour length scale λ_c, and memory decay γ, achieved a R^2 score of 0.011, 0.001, and 0.733, respectively. These results suggest that the memory decay parameter is the most influential in correctly modelling human errors in the colour reproduction task. Therefore, we hypothesise that misremembering the colours of previously seen objects is more due to memory capacity than spatial and feature uncertainty. The significant improvement of our SSP-decay model compared to the SSP-base model further supports this hypothesis.

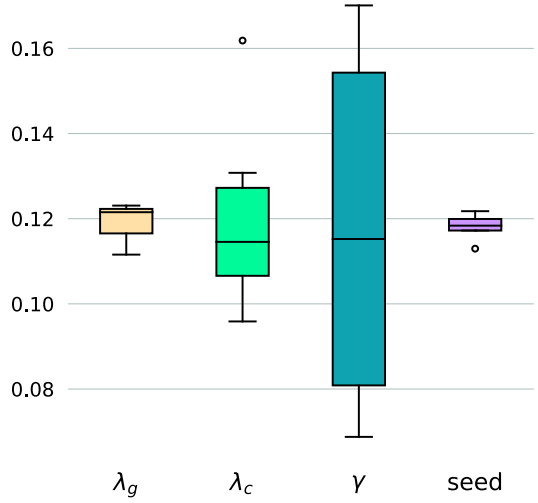


Figure 3: Parameter sensitivity for **SSP-decay** with distribution response. Effect of grid length scale λ_g, colour length scale λ_c, memory decay γ, and random seeds on average KL divergence for target and non-target distribution.

To highlight the interpretability of our hyperparameters, we show their effect in spatial and feature dimensions. Figure 4 visualises an example trial of Experiment 1. From left to right, the hyperparameters match the optimised parameters for the base models and the decay models with both response types; however, we set $\gamma = 1$ for visualisation purposes. On the top (a), we see the effect of the grid length scale parameter

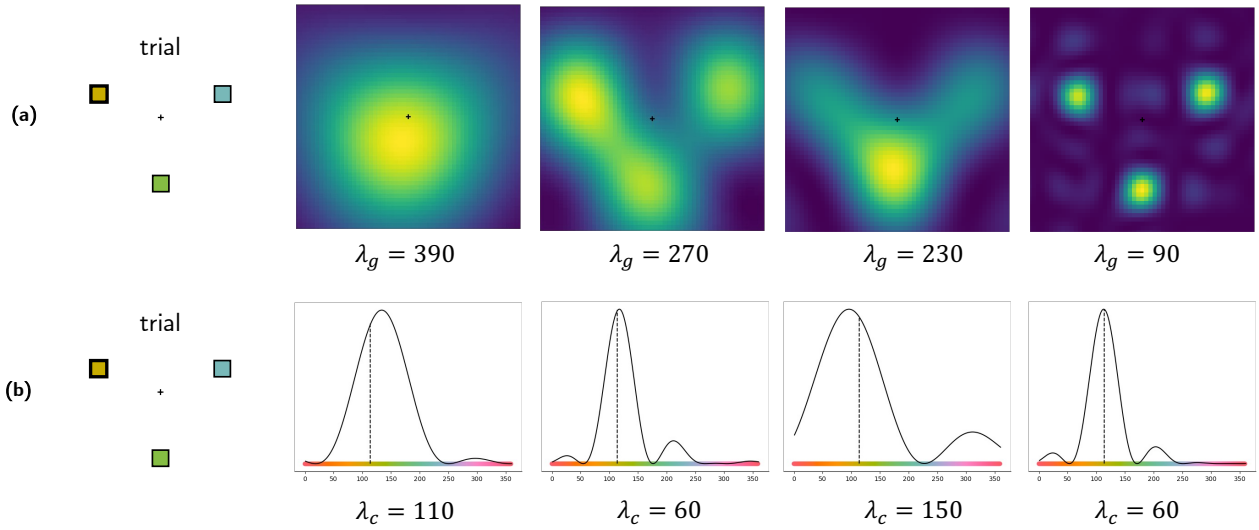


Figure 4: Parameter visualization. In **(a)**, we see the effect of different length scales in the spatial dimension λ_g, and in **(b)**, the effect of different colour length scales λ_c on the decoding of the probed object (yellow).

λ_g. Here we probe all given colours via unbinding and overlay the resulting spatial similarity maps (brighter colours indicate higher similarity). A large length scale like $\lambda_g = 390$ introduces high spatial uncertainty, effectively merging all object locations into a single point. In contrast, $\lambda_g = 90$ allows us to distinguish all objects in memory. Similarly, in Figure 4 (b), we plot the colour similarity distribution when probing with the target location (yellow). A high colour length scale $\lambda_c = 150$ yields a wide similarity distribution, i.e. models an increased uncertainty in the feature domain. In contrast, the lower length scale $\lambda_c = 60$ has much less variance and is correctly centred around the target colour (dashed line).

Discussion

The base models with no memory decay require high spatial and feature uncertainty to capture the human data, and both parameters must be even higher for the maximum response models. The distribution response model introduces additional variance and also better predicts human errors. Adding the decay parameter γ significantly improved model fit and allowed for lower spatial and feature uncertainties. Our best model – SSP-decay (dist) – exhibits very high spatial certainty and a smaller variance in feature dimension than other models. In general, we find that a mixture of maximum and response distribution would fit the data best. In detail, the maximum response for set sizes below four and the distribution response for higher set sizes are taken. We find this trade-off worth investigating in future work and are interested to see whether this hints towards differing decision paradigms.

We were not able to quantitatively compare our results to the IM (Oberauer & Lin, 2017) as they used different evaluation criteria, and we were unable to replicate their fine-tuned, participant-specific parameters. However, we found our results to be qualitatively on par, and we want to point out the key advantages of our modelling approach. Our proposed method does not require fixed kernel shapes and generalises across different experiments without additional fine-tuning or subject-specific fine-tuning in general. While Oberauer & Lin (2017) had a decay parameter in their IM-DR model, they report a significant decrease in model fit for the cueing experiments. In all our models, we observed *increasing* model fit on Experiments 2 and 3 without performing any hyper-parameter fine-tuning. This highlights the generalisability of our proposed VWM model.

More generally, our method is embedded in the SPA cognitive modelling framework, has been shown to easily integrate a cognitive model of attention (*e.g.* EMMA (Salvucci, 2001)), and allows for easy implementation in spiking neural networks, which we intend to provide in future work. Our model only requires three parameters, in contrast to the six parameters introduced by the IM. Our parameters are also highly interpretable and not too sensitive, i.e., the base model already captures the human phenomena relatively well.

Additionally, our method can naturally represent conjunctions of features through binding without increasing vector dimension. Hence, the size of conjunction representations grows slower than the linear growth required by Oberauer & Lin (2017) or the exponential growth required by conjunctive coding (Schneegans & Bays, 2017).

The integration of the cognitive attention model EMMA improved our model fit only slightly, but we believe it will be beneficial for more complex stimuli, where the scan path and time estimates are more meaningful. In future work, it will also be interesting to integrate more complex models of attention, e.g., PAAV (Nyamsuren & Taatgen, 2013). We also plan to extend our method to more complex natural scenes, as proposed by Bates, Alvarez, & Gershman (2023).

Acknowledgements

Anna Penzkofer was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2075 – 390740016. Michael Furlong was funded in part by CFI and OIT infrastructure funding, NSERC Discovery grant 261453, NUCC NRC File A-0028850, AFOSR grant FA9550-17-1-0026, and an Intel Neuromorphic Research Community Grant. Chris Eliasmith was funded in part by the Canada Research Chairs program, CFI and OIT infrastructure funding, NSERC Discovery grant 261453, NUCC NRC File A-0028850, and AFOSR grant FA9550-17-1-0026.

References

Alvarez, G., & Cavanagh, P. (2004). The Capacity of Visual Short-Term Memory is Set Both by Visual Information Load and by Number of Objects. *Psychological Science*, *15*. doi: 10.1111/j.0963-7214.2004.01502006.x

Anderson, J. R., Matessa, M., & Lebiere, C. (1997). ACT-R: A Theory of Higher Level Cognition and Its Relation to Visual Attention. *Human–Computer Interaction*, *12*(4). doi: 10.1207/s15327051hci1204_5

Bates, C. J., Alvarez, G. A., & Gershman, S. J. (2023). *Scaling models of visual working memory to natural images.* bioRxiv. doi: 10.1101/2023.03.17.533050

Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science (New York, N.Y.)*, *321*(5890). doi: 10.1126/science.1158023

Bays, P. M., Schneegans, S., Ma, W. J., & Brady, T. F. (2024). Representation and computation in visual working memory. *Nature Human Behaviour*, *8*(6). doi: 10.1038/s41562-024-01871-2

Chai, W. J., Abd Hamid, A. I., & Abdullah, J. M. (2018). Working Memory From the Psychological and Neurosciences Perspectives: A Review. *Frontiers in Psychology*, *9*. doi: 10.3389/fpsyg.2018.00401

Choo, F.-X., & Eliasmith, C. (2010). A spiking neuron model of serial-order recall. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 32).

Dumont, N. S.-Y., & Eliasmith, C. (2020). Accurate representation for spatial cognition using grid cells. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 42).

Eliasmith, C. (2013). *How to Build a Brain: A Neural Architecture for Biological Cognition*. Oxford University Press. doi: 10.1093/acprof:oso/9780199794546.001.0001

Eliasmith, C., Stewart, T. C., Choo, X., Bekolay, T., DeWolf, T., Tang, Y., & Rasmussen, D. (2012). A large-scale model of the functioning brain. *science*, *338*(6111), 1202–1205.

Furlong, P. M., & Eliasmith, C. (2023). Modelling neural probabilistic computation using vector symbolic architectures. *Cognitive Neurodynamics*. doi: 10.1007/s11571-023-10031-7

Gayler, R. W. (2004). *Vector Symbolic Architectures answer Jackendoff's challenges for cognitive neuroscience.* arXiv. doi: 10.48550/arXiv.cs/0412059

Gosmann, J., & Eliasmith, C. (2015). A Spiking Neural Model of the n-Back Task. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 37).

Komer, B., Stewart, T., Voelker, A., & Eliasmith, C. (2019). A neural representation of continuous space using fractional binding. In *Annual Meeting of the Cognitive Science Society.*

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657). doi: 10.1038/36846

Nyamsuren, E., & Taatgen, N. A. (2013). Pre-attentive and attentive vision module. *Cognitive Systems Research*, *24*. doi: 10.1016/j.cogsys.2012.12.010

Oberauer, K., & Lin, H.-Y. (2017). An interference model of visual working memory. *Psychological Review*, *124*(1). doi: 10.1037/rev0000044

Plate, T. A. (2003). *Holographic Reduced Representation: Distributed Representation for Cognitive Structures*. Center for the Study of Language and Information.

Salvucci, D. D. (2001). An integrated model of eye movements and visual encoding. *Cognitive Systems Research*, *1*(4). doi: 10.1016/S1389-0417(00)00015-2

Schneegans, S., & Bays, P. M. (2017). Neural Architecture for Feature Binding in Visual Working Memory. *Journal of Neuroscience*, *37*(14). doi: 10.1523/JNEUROSCI.3493-16.2017

Schöner, G., & Spencer, J. P. (2016). *Dynamic thinking: A primer on dynamic field theory*. Oxford University Press.

van den Berg, R., Shin, H., Chou, W.-C., George, R., & Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences*, *109*(22). doi: 10.1073/pnas.1117465109

Verghese, P. (2001). Visual search and attention: A signal detection theory approach. *Neuron*, *31*(4), 523–535.

Wilken, P., & Ma, W. J. (2004). A detection theory account of visual short-term memory for color. *Journal of Vision*, *4*(8). doi: 10.1167/4.8.150

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*(7192). doi: 10.1038/nature06860