Neural and computational arguments for memory as a compressed supported timeline

Zoran Tiganj (zoran.tiganj@gmail.com)
Karthik H. Shankar (kshankar79@gmail.com)
Marc W. Howard (marcwhoward777@gmail.com)
Center for Memory and Brain
Department of Psychological and Brain Sciences
Boston University, Boston, MA 02215

Abstract

It is well known that, all things being equal, the accuracy of mammalian timing and memory decays gradually with the passage of time. The gradual decay of temporal accuracy is also observed in single-unit neural recordings. Here we review recent modeling work describing a specific mechanism for timing and memory and relevant neural data. The model describes a neural mechanism that can give rise to a logarithmically compressed representation of the recent past. We examine the specific predictions of the model, in particular that the elapse of time is represented by sequentially activated cells which fire for a circumscribed period of time. Such cells, called *time cells*, have been observed in neural recordings from several brain regions in multiple species. As predicted by the model, the cells show accuracy that decreases with time.

Keywords: scale-invariance, memory, interval timing, time cells.

Introduction

Behavioral experiments on humans and other animals have demonstrated that the accuracy in estimating the duration of a time interval decays gradually with the interval duration itself. More specifically, the variability of the response is proportional to the interval duration (Rakitin et al., 1998; Ivry & Hazeltine, 1995). For instance, in interval timing the response distributions appear to be scale-invariant: distributions corresponding to different interval durations overlap when linearly scaled (Roberts, 1981; Smith, 1968). Furthermore, animal literature suggests that in conditioning paradigms, the number of trials needed for an animal to learn the association between conditioned and unconditioned stimuli scales with a ratio of the reinforcement latency and intertrial interval (Gallistel & Gibbon, 2000), indicating again the scale-invariance in the animals' behavior (Balsam & Gallistel, 2009; Shankar & Howard, 2012; Gibbon, 1977).

In addition to timing, scale-invariance has been argued to be one of the key properties of memory. Gradual decay of memory without a characteristic scale has been observed in a number of behavioral experiments (Anderson & Schooler, 1991; Chater & Brown, 2008; Wixted & Ebbesen, 1991) and it was often refereed to as *power-law of forgetting*. For example, Donkin and Nosofsky (2012) reported that in item recognition task the strength of the memory was decaying as a power law function of the lag between studied items and a test probe. It has been argued that scale-invariance should be thought of as universal law of cognition (Chater & Brown, 1999). A number of cognitive models have been constructed to account for these properties (Brown, Neath, & Chater, 2007; Howard, Shankar, Aue, & Criss, 2015; Donkin & Nosofsky, 2012).

Neural mechanisms that could support the scale-invariance of time and memory are still unclear. It has been argued that working memory is represented with persistent neural activity observed in different areas of the prefrontal cortex (PFC) during for instance a delayed match to sample task (P. S. Goldman-Rakic, 1991; P. Goldman-Rakic, 1995). Even though such persistent activity can account well for the demands of a particular task, it is not clear how it could account for a gradual decay of the memory representation. More recent studies have found that in some behavioral tasks a subset of neurons activates sequentially, tiling the task relevant interval, typically lasting for several seconds (see e.g. Pastalkova, Itskov, Amarasingham, and Buzsaki (2008); MacDonald, Lepage, Eden, and Eichenbaum (2011) for the first reports of such activity). These neurons that fire sequentially, each during a circumscribed period of time, are called *time cells* (Howard & Eichenbaum, 2015; Eichenbaum, 2013). It has been argued that time cells play important role in timing and memory (MacDonald, Fortin, Sakata, & Meck, 2014; Howard et al., 2014; Eichenbaum, 2014).

Time cells provide a direct readout of when the delay interval has started: there is no need for population decoding in a classical sense (Murray et al., 2016; Stokes, 2015). This is because time cells activate sequentially effectively providing temporal basis functions and constituting an internal timeline. As we will discuss later, this timeline is compressed, such that the temporal resolution gradually decays with the elapsed time, just as expected given the behavioral findings on memory and timing we mentioned above. It is unclear, however, what are the neural mechanisms that could give rise to such a compressed timeline.

Here we utilize a computational model for compressed scale-invariant dynamical memory representation introduced in Shankar and Howard (2012). We compare specific predictions of the model with the neural recordings of sequentially activated time cells. The model provides a unique solution to constructing a scale-invariant memory representation (Shankar, 2015). The model has been used to account for results of various timing and memory experiments including judgment of recency and serial scanning in long and short term memory (Howard, Shankar, Aue, & Criss, 2015). Here

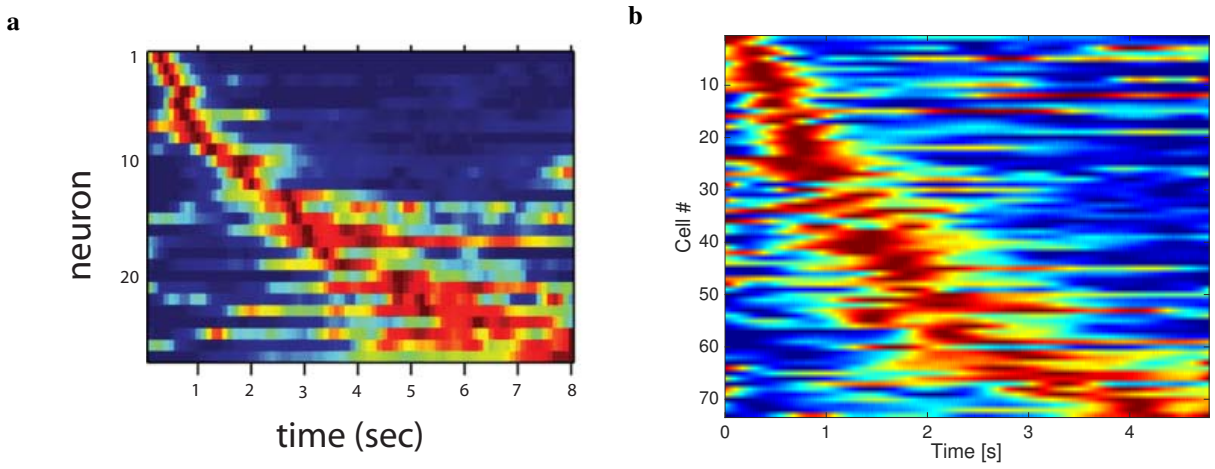

Figure 1: Examples of sequentially activated neurons from tetrode recordings in rat hippocampus (plot **a.**) and PFC (plot **b.**). On both plots each row on the heatplot corresponds to a single cell and displays normalized firing rate averaged across trials during a delay interval of a behavioral experiment. Red corresponds to high firing rate, while blue corresponds to low firing rate. The cells are sorted with respect to their peak time. Despite the fact that both recordings are done during a rather different behavioral experiment, they both show similar qualitative properties. In particular we point to two features related to the temporal accuracy: 1) time fields later in the delay are more broad than time fields earlier in the delay (the central ridge is widening as the peak moves to the right); 2) peak times of the time cells are not evenly distributed across the delay, with later time periods represented by fewer cells than early time periods (this is apparent from the curvature of the central ridge; a uniform distribution of time fields would manifest as a straight line). **a.** Hippocampal CA1 neurons recorded during object-delay-odor sequence task (reprint from MacDonald et al. (2011)). In order to obtain a reward the animals had to memorize the identity of the stimulus during the delay interval and match it to the appropriate odor. **b.** PFC neurons recorded during a temporal discrimination task (reprint from Tiganj et al. (2016), original data first reported in (Kim et al., 2013)). In order to obtain a reward the animals had to estimate whether the presented delay interval was larger than some baseline duration and make a left or right turn accordingly following the delay.

we focus on the neural side, looking into specific predictions about individual neural activity that can be derived from the model.

Sequential activation as a neural correlate of timing and memory

Starting with a report from Pastalkova et al. (2008), a number of studies have reported sequential neural activation in different timing and memory tasks from different brain regions: hippocampus (MacDonald et al., 2011; Salz et al., 2016; Gill, Mizumori, & Smith, 2011; Kraus, Robinson, White, Eichenbaum, & Hasselmo, 2013; MacDonald, Carrow, Place, & Eichenbaum, 2013; Modi, Ashesh, & Bhalla, 2014; Naya & Suzuki, 2011), PFC (Tiganj et al., 2016) and the striatum (Mello, Soares, & Paton, 2015; Adler et al., 2012) in a variety of behavioral tasks. This activity has been hypothesized to be a neural basis for representation of memory and elapsed time in a gradually decaying fashion (Howard, Shankar, Aue, & Criss, 2015; Howard & Eichenbaum, 2015; Eichenbaum, 2014, 2013). The studies were done on different animals, including rats (MacDonald et al., 2011; Salz et al., 2016; Gill et al., 2011; Kraus et al., 2013; MacDonald et al., 2013; Pastalkova et al., 2008; Mello et al., 2015; Tiganj et al., 2016), mice (Modi et al., 2014) and monkeys (Naya & Suzuki, 2011; Adler et al., 2012). Even though the majority of studies used tetrode recordings, Modi et al. (2014) used two-photon calcium imaging minimizing the probability that the results were observed due to some sort of recording artifact. Most of the studies were done on animals that were allowed to move, but some were done on head-fixed animals (MacDonald et al., 2013; Modi et al., 2014; Naya & Suzuki, 2011; Adler et al., 2012) confirming that the results were not coming from position-related artifacts.

It is worth noting that sequentially activated time cells were observed in these studies despite the different cognitive demands on the animals, which included temporal discrimination (e.g. Tiganj et al. (2016)) or memory demands (e.g. Salz et al. (2016); MacDonald et al. (2011)). The duration of the intervals where such cells were measured was ranging from a couple of seconds up to 60 s (Mello et al., 2015).

Several studies have observed decreasing temporal accuracy as a function of delay, due to spread in time field width (Howard et al., 2014; MacDonald et al., 2011; Mello et al., 2015; Adler et al., 2012; Kraus et al., 2013; Salz et al., 2016; Tiganj et al., 2016) and/or due to a non-uniform distribution of time fields (Kraus et al., 2013; Salz et al., 2016; Mello et al., 2015; Tiganj et al., 2016). Two examples of neural representation with decreasing temporal accuracy are provided in Figure 1.

Computational model for compressed scale-invariant dynamical memory representation

The computational model reviewed here was initially introduced in (Shankar & Howard, 2012). It consists of a two-layer feedforward neural network with analytically derived weights. Here we briefly describe the model and then focus on its predictions regarding neural activity. Notice that below

we define the model as a model of memory, as it was initially introduced in (Shankar & Howard, 2012). Its application in timing is restricted to the stimulus that initiates the delay interval, as the only stimulus that needs to be remembered.

We first define an input vector $\mathbf{f}$ consisting of N elements such that each of its elements corresponds to a unique stimulus. Thus observing for example stimulus A makes an element in $\mathbf{f}$ that corresponds to stimulus A, f_A, equal to one for the time A is presented and zero otherwise. Each element of the input vector $\mathbf{f}$ has a two-layer dynamical compressed memory representation. The first layer of the network implements an approximation of an integral transform of the input (Laplace transform, but as a function of a real rather than a complex variable). This means that nodes in the first layer, $\mathbf{F}(t,\overset{*}{t})$, act as leaky integrators (first order low-pass filters) with a spectrum of time constant defined with $k/\overset{*}{t}$, where k is positive integer (Figure 2):

$$\frac{\mathbf{F}(t,\overset{*}{t})}{dt} = -\frac{k}{\overset{*}{t}}\mathbf{F}(t,\overset{*}{t}) + \mathbf{f}(t). \tag{1}$$

Leaky integrators project to the second layer, $\tilde{\mathbf{f}}$, through fixed weights that implement an approximation of the inverse of the transform by applying a k^{th} order derivative with respect to $k/\overset{*}{t}$, denoted as $\mathbf{F}^{(k)}(t,\overset{*}{t})$ (the inverse is derived based on Post's inversion formula (Post, 1930), see Shankar and Howard (2012, 2013) for further details on the derivation):

$$\tilde{\mathbf{f}}(t,\overset{*}{t}) = C_k \left(\frac{k}{\overset{*}{t}}\right)^{k+1} \mathbf{F}^{(k)}(t,\overset{*}{t}), \tag{2}$$

where C_k is a constant that depends only on k. The cells in the second layer constitute a dynamical memory representation of the input signal. To understand the properties of the memory representation we consider an impulse response of a cell in $\tilde{\mathbf{f}}$. For $f_A(\tau) = \delta(\tau = 0)$ the corresponding activation of the cells in the second layer is:

$$\tilde{f}_A(t,\overset{*}{t}) = C_k \frac{1}{\overset{*}{t}} \left(\frac{t}{\overset{*}{t}}\right)^{k} e^{-k\frac{t}{\overset{*}{t}}}, \tag{3}$$

where C_k here is a different constant that depends only on k. The activity of each node in $\tilde{f}_A(t,\overset{*}{t})$ is the product of an increasing power term $\left(\frac{t}{\overset{*}{t}}\right)^k$ and a decreasing exponential term $e^{-k\frac{t}{\overset{*}{t}}}$. Consequently, each node in $\tilde{f}_A(t,\overset{*}{t})$ has a peak that corresponds to the $\overset{*}{t}$ value of that node: $\frac{d\tilde{f}_A(t,\overset{*}{t})}{dt} = 0 \Rightarrow t = \overset{*}{t}$. Thus, following a transient input, cells in $\tilde{f}_A$ activate sequentially in time constituting a dynamical memory representation of the input A (Figure 3).

This memory representation has perfect accuracy in the limit when $k \to \infty$. In a realistic biological or artificial neural network, where k is finite and $\overset{*}{t}$ is a discrete variable sup-

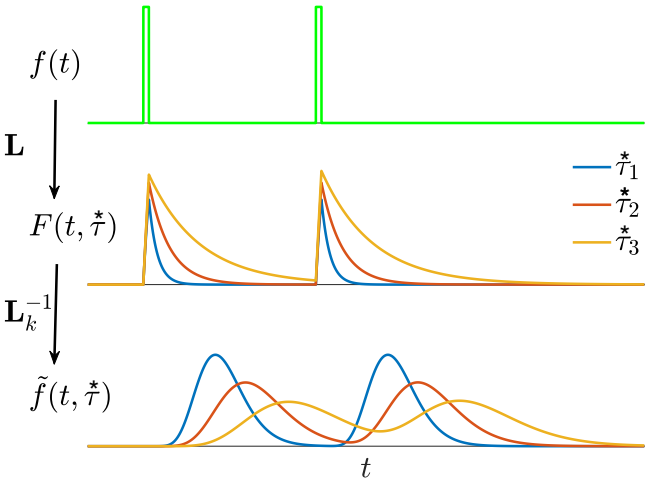

Figure 2: Constructing a scale-invariant compressed memory representation through an integral transform and its inverse. A transient input stimulus $\mathbf{f}(t)$ (top row) is presented twice and feeds into a layer of leaky integrators $\mathbf{F}(t,\overset{*}{t})$ with a spectrum of time constants $\overset{*}{t}$ constituting a discrete approximation of an integral transform (middle row). The transform is denoted as $\mathbf{L}$ since it is equivalent to the real part of the Laplace transform. Only three nodes in $\mathbf{F}(t,\overset{*}{t})$ are shown. Each leaky integrator is characterized with its time constant, $\overset{*}{t}$. $\mathbf{F}$ projects onto $\tilde{\mathbf{f}}(t,\overset{*}{t})$ through a set of weights defined with the operator denoted as $\mathbf{L}_k^{-1}$ which implements an approximation of the inverse of the Laplace transform. Nodes in $\tilde{\mathbf{f}}(t,\overset{*}{t})$ activate sequentially following the stimulus presentation creating a memory representation. The width of the activation of each node scales with the peak time determined by the corresponding $\overset{*}{t}$, making the memory scale-invariant. Logarithmic spacing of the $\overset{*}{t}$ assures that the memory representation is compressed.

ported with a limited number of nodes, the memory representation becomes an approximation of the past. The approximation is scale-invariant (Figure 4) since the width of the activation of each node scales with the peak time (this is scale-invariant since rescaling the temporal axis rescales the width of the activation by the same amount). In other words, the accuracy of the memory representation decreases with the elapse of time since the stimulus presentation. With appropriately distributed $\overset{*}{t}$ the representation can be made logarithmically compressed.

To establish biological plausibility of the model we have shown that leaky integrators with a spectrum of time constants are biologically realistic (Tiganj, Hasselmo, & Howard, 2015; Tiganj, Shankar, & Howard, 2013). In addition, taking derivatives with respect to $k/\overset{*}{t}$ amounts to lateral inhibition, making it biologically plausible as well (Howard et al., 2014). To implement the derivative it is required that each neurons

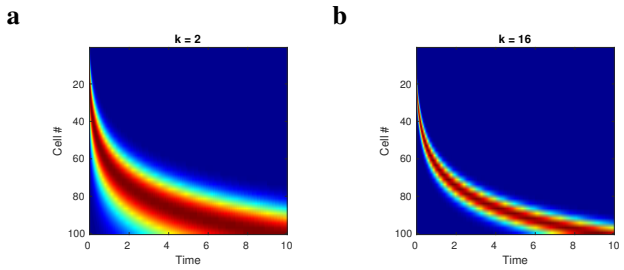

Figure 3: Activity of the cells in the compressed memory representation generated by the model. Analogous to the heatmaps in Figure 1, each row corresponds to a single cell and displays its normalized activity across time. The cells are sorted with respect to the peak time defined by their value of $\overset{*}{t}$. The two features observed in Figure 1 are fully captured by the model: the time fields later in the delay were more broad than the time fields earlier in the delay and the density of time fields decreased as a function of time ($\overset{*}{t}$ was logartihmically spaced). This illustrates that the model can indeed account for the firing dynamics of the sequentially activated time cells that form a compressed representation of time. The two plots, **a** and **b**, show the activity of the cell ensemble for two different values of parameter k. Increasing k makes the firing fields more narrow and the memory representation more precise. Notice that, from the biological perspective, larger k is more difficult to obtain, since it requires higher order of derivative with respect to $k/\overset{*}{t}$. This requires broader connectivity between the two layers.

of the first layer only projects to the k neighboring neurons of the second layer. The connectivity pattern is the same across the entire projection, since it always implements a derivative with respect to $k/\overset{*}{t}$. In addition, qualitative alignment of the model with the sequential neural activity as shown in Figure 3 further supports its biological plausibility.

Discussion

We reviewed the predictions from a computational model for compressed scale-invariant memory representation and compared them to the results from recently-published neural recordings. The model maintains a dynamical representation of the recent past through a set of sequentially activated neurons. Such sequential activation appears qualitatively similar to the data published in multiple studies over the past several years including different regions of the brain including the hippocampus, PFC and striatum.

Several of the studies align with the model exhibiting compressed memory representation. In particular, the width of the time fields increased with the peak time and more cells had time fields earlier than later in the delay interval (notice the common trend in the plots in Figure 1 and Figure 3). These findings suggest that the model can indeed account for the neural representation of the elapsed time.

The model makes specific prediction on the scale-invariance of the memory representation which was inspired by the behavioral experiments on timing and memory. Existing neural data were thus far not sufficient to explicitly test that prediction. However, the qualitative observations made here are consistent with the scale-invaraince prediction, but they are not sufficient to quantitatively verify it.

In addition to the model described here, several other computational models predict the qualitative properties found in the data. The common aspect of most of such models is the functional form that gives rise to time fields: as in the model described here, the activity increase is governed by a power-law and then later attenuated by a damping exponential. In particular, Grossberg and Schmajuk (1989); De Vries and Principe (1992); Machado (1997) propose different mechanistic solutions for achieving such form. However, unlike in the model described here, rescaling the time axis in these models would change the functional form of the representation. Others (for instance Tank and Hopfield (1987); Ludvig, Sutton, and Kehoe (2012)) directly used the functional form that provides spreading temporal basis functions as seen here.

Experimental data allowed us thus far to verify some of the predictions computational models make regarding the compressed representation of time. However, the model described here makes specific predictions regarding how memory is maintained in general. Here we assumed that the stimulus that marks the onset of the delay interval is the only one that has the memory representation. The model is designed to capture a variety of stimuli and maintain an independent compressed memory representation for each of them. In fact, associations between the independent representations allowed us to test the model on a variety of memory tasks (Howard, Shankar, Aue, & Criss, 2015). It is to be tested whether the neural representation indeed supports such independent, stimulus specific compressed memory representations (see Tiganj, Cromer, Roy, Miller, and Howard (2017) for recent evidence of this).

Maintaining temporal information through sequential activation has a critical computational property in that it provides a direct readout of the elapsed time. Notice that cells in the first layer of the model (leaky integrators) contain the same amount of temporal information as the cells in the second layer (sequentially activated neurons). Thus one could apply population decoding techniques and extract the temporal information from the first layer directly. In fact, this is exactly what the inverse transform is doing. However, instead of training a classifier, which would be a common decoding procedure, it provides a simple form of linear readout using a mechanism analogous to lateral inhibition, which is known to exist in the nervous system. An additional advantage of having such a mechanism is that it provides access to the real-value Laplace domain, where computations that are otherwise hard to achieve in a neural network become straightforward. These in particular include addition and subtraction of probability distributions as well as temporal translation (Howard, Shankar, & Tiganj, 2015; Shankar, Singh, & Howard, 2016).

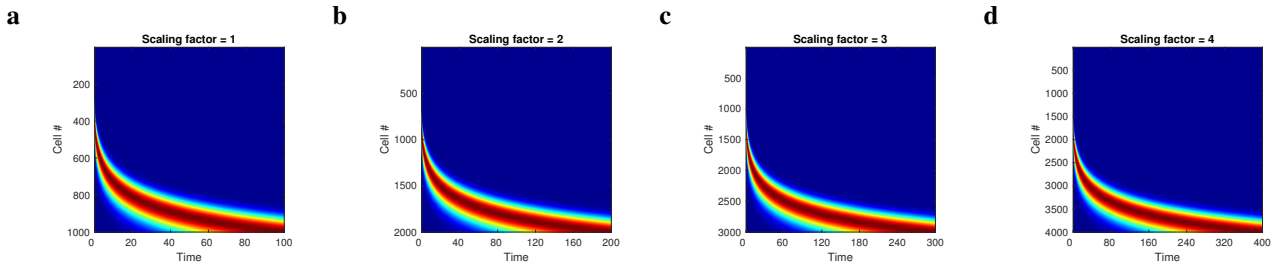

Figure 4: Illustration of scale-invariance in the compressed memory representation generated by the model. Scaling the number of cells and the temporal duration by the same factor results in identical memory representation (plots **a.** to **d.** appear identical despite the fact that both x and y axes are rescaled on each plot). This property follows from Equations (3) since $\overset{*}{t}$ and t appear only as a ratio (except for the scaling factor in front that does not influence the functional form). Scale-invariance is consistent with behavioral experiments, but it remains unclear whether neural data exhibits this property as well, even though the results shown in Figure 1 are consistent with scale-invariance.

Conclusion

We showed that a computational model for constructing compressed dynamical representations of the recent past aligns well with recent neural data showing sequential neural activation. The sequential activation constitutes a compressed supported timeline, providing a mechanism for representing the elapse of time and potentially a mechanism for maintaining a dynamical memory representation.

Acknowledgments

This work was supported by NIBIB R01EB022864, NIMH R01MH112169 and MURI N00014-16-1-2832.

References

Adler, A., Katabi, S., Finkes, I., Israel, Z., Prut, Y., & Bergman, H. (2012). Temporal convergence of dynamic cell assemblies in the striato-pallidal network. *Journal of Neuroscience*, *32*(7), 2473-84. doi: 10.1523/JNEUROSCI.4830-11.2012

Anderson, J., & Schooler, L. (1991). Reflections of the environment in memory. *Psychological science*, *2*(6), 396–408.

Balsam, P. D., & Gallistel, C. R. (2009). Temporal maps and informativeness in associative learning. *Trends in Neuroscience*, *32*(2), 73–78.

Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review*, *114*(3), 539-76.

Chater, N., & Brown, G. D. (1999). Scale-invariance as a unifying psychological principle. *Cognition*, *69*(3), B17–B24.

Chater, N., & Brown, G. D. A. (2008). From universal laws of cognition to specific cognitive models. *Cognitive Science*, *32*(1), 36-67. doi: 10.1080/03640210701801941

De Vries, B., & Principe, J. C. (1992). The gamma modela new neural model for temporal processing. *Neural networks*, *5*(4), 565–576.

Donkin, C., & Nosofsky, R. M. (2012). A power-law model of psychological memory strength in short- and long-term recognition. *Psychological Science*. doi: 10.1177/0956797611430961

Eichenbaum, H. (2013). Memory on time. *Trends in Cognitive Sciences*, *17*(2), 81-8. doi: 10.1016/j.tics.2012.12.007

Eichenbaum, H. (2014). Time cells in the hippocampus: a new dimension for mapping memories. *Nature Reviews Neuroscience*, *15*(11), 732–744.

Gallistel, C. R., & Gibbon, J. (2000). Time, rate, and conditioning. *Psychological Review*, *107*(2), 289-344.

Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychological Review*, *84*(3), 279-325.

Gill, P. R., Mizumori, S. J. Y., & Smith, D. M. (2011). Hippocampal episode fields develop with learning. *Hippocampus*, *21*(11), 1240-9. doi: 10.1002/hipo.20832

Goldman-Rakic, P. (1995). Cellular basis of working memory. *Neuron*, *14*, 477-85.

Goldman-Rakic, P. S. (1991). Cellular and circuit basis of working memory in prefrontal cortex of nonhuman primates. *Progress in brain research*, *85*, 325–336.

Grossberg, S., & Schmajuk, N. (1989). Neural dynamics of adaptive timing and temporal discrimination during associative learning. *Neural Networks*, *2*(2), 79–102.

Howard, M. W., & Eichenbaum, H. (2015). Time and space in the hippocampus. *Brain research*, *1621*, 345–354.

Howard, M. W., MacDonald, C. J., Tiganj, Z., Shankar, K. H., Du, Q., Hasselmo, M. E., & Eichenbaum, H. (2014). A unified mathematical framework for coding time, space, and sequences in the hippocampal region. *Journal of Neuroscience*, *34*(13), 4692-707. doi: 10.1523/JNEUROSCI.5808-12.2014

Howard, M. W., Shankar, K. H., Aue, W., & Criss, A. H. (2015). A distributed representation of internal time. *Psychological Review*, *122*(1), 24–53. doi: 10.1037/a0037840

Howard, M. W., Shankar, K. H., & Tiganj, Z. (2015). Efficient neural computation in the laplace domain. In *Proceedings of the 2015th international conference on cognitive computation: Integrating neural and symbolic approaches-volume 1583* (pp. 61–68).

Ivry, R. B., & Hazeltine, R. E. (1995, Feb). Perception and production of temporal intervals across a range of dura-

tions: evidence for a common timing mechanism. *J Exp Psychol Hum Percept Perform*, *21*(1), 3-18.

Kim, J., Ghim, J.-W., Lee, J. H., & Jung, M. W. (2013). Neural correlates of interval timing in rodent prefrontal cortex. *Journal of Neuroscience*, *33*(34), 13834-47. doi: 10.1523/JNEUROSCI.1443-13.2013

Kraus, B. J., Robinson, R. J., 2nd, White, J. A., Eichenbaum, H., & Hasselmo, M. E. (2013). Hippocampal "time cells": time versus path integration. *Neuron*, *78*(6), 1090-101. doi: 10.1016/j.neuron.2013.04.015

Ludvig, E. A., Sutton, R. S., & Kehoe, E. J. (2012). Evaluating the TD model of classical conditioning. *Learning & Behavior*, *40*(3), 305–319.

MacDonald, C. J., Carrow, S., Place, R., & Eichenbaum, H. (2013). Distinct hippocampal time cell sequences represent odor memories immobilized rats. *Journal of Neuroscience*, *33*(36), 14607–14616.

MacDonald, C. J., Fortin, N. J., Sakata, S., & Meck, W. H. (2014). Retrospective and prospective views on the role of the hippocampus in interval timing and memory for elapsed time. *Timing & Time Perception*, *2*(1), 51–61.

MacDonald, C. J., Lepage, K. Q., Eden, U. T., & Eichenbaum, H. (2011). Hippocampal "time cells" bridge the gap in memory for discontiguous events. *Neuron*, *71*, 737-749.

Machado, A. (1997). Learning the temporal dynamics of behavior. *Psychological review*, *104*(2), 241.

Mello, G. B. M., Soares, S., & Paton, J. J. (2015). A Scalable Population Code for Time in the Striatum. *Current Biology*, *25*(9), 1113–1122.

Modi, N. M., Ashesh, D. K., & Bhalla, S. U. (2014). CA1 cell activity sequences emerge after reorganization of network correlation structure during associative learning. *eLife*, *3*(0). Retrieved from `http://dx.doi.org/10.7554/eLife.01982` doi: 10.7554/eLife.01982

Murray, J. D., Bernacchia, A., Roy, N. A., Constantinidis, C., Romo, R., & Wang, X.-J. (2016). Stable population coding for working memory coexists with heterogeneous neural dynamics in prefrontal cortex. *Proceedings of the National Academy of Sciences*, 201619449.

Naya, Y., & Suzuki, W. (2011). Integrating what and when across the primate medial temporal lobe. *Science*, *333*(6043), 773-776.

Pastalkova, E., Itskov, V., Amarasingham, A., & Buzsaki, G. (2008). Internally generated cell assembly sequences in the rat hippocampus. *Science*, *321*(5894), 1322-7.

Post, E. (1930). Generalized differentiation. *Transactions of the American Mathematical Society*, *32*, 723-781.

Rakitin, B. C., Gibbon, J., Penny, T. B., Malapani, C., Hinton, S. C., & Meck, W. H. (1998). Scalar expectancy theory and peak-interval timing in humans. *Journal of Experimental Psychology: Animal Behavior Processes*, *24*, 15-33.

Roberts, S. (1981). Isolation of an internal clock. *Journal of Experimental Psychology: Animal Behavior Processes*, *7*, 242-268.

Salz, D. M., Tiganj, Z., Khasnabish, S., Kohley, A., Sheehan, D., Howard, M. W., & Eichenbaum, H. (2016). Time cells in hippocampal area ca3. *The Journal of Neuroscience*, *36*(28), 7476–7484.

Shankar, K. H. (2015). Generic construction of scale-invariantly coarse grained memory. In *Australasian conference on artificial life and computational intelligence* (pp. 175–184).

Shankar, K. H., & Howard, M. W. (2012). A scale-invariant representation of time. *Neural Computation*, *24*, 134-193.

Shankar, K. H., & Howard, M. W. (2013). Optimally fuzzy scale-free memory. *Journal of Machine Learning Research*, *14*, 3753-3780.

Shankar, K. H., Singh, I., & Howard, M. W. (2016). Neural mechanism to simulate a scale-invariant future. *Neural Computation*, *28*(12).

Smith, M. C. (1968). CS-US interval and US intensity in classical conditioning of rabbit's nictitating membrane response. *Journal of Comparative and Physiological Psychology*, *3*, 679-687.

Stokes, M. G. (2015). 'activity-silent' working memory in prefrontal cortex: a dynamic coding framework. *Trends in Cognitive Sciences*, *19*(7), 394–405.

Tank, D., & Hopfield, J. (1987). Neural computation by concentrating information in time. *Proceedings of the National Academy of Sciences*, *84*(7), 1896–1900.

Tiganj, Z., Cromer, J. A., Roy, J. E., Miller, E. K., & Howard, M. W. (2017). Compressed timeline of recent experience in monkey lpfc. *bioRxiv*, 126219.

Tiganj, Z., Hasselmo, M. E., & Howard, M. W. (2015). A simple biophysically plausible model for long time constants in single neurons. *Hippocampus*, *25*(1), 27-37.

Tiganj, Z., Kim, J., Jung, M. W., & Howard, M. W. (2016). Sequential firing codes for time in rodent mPFC. *Cerebral Cortex*(1-9). doi: 10.1093/cercor/bhw336

Tiganj, Z., Shankar, K. H., & Howard, M. W. (2013). Encoding the laplace transform of stimulus history using mechanisms for persistent firing. *BMC Neuroscience*, *14*(Suppl 1), P356.

Wixted, J. T., & Ebbesen, E. B. (1991). On the form of forgetting. *Psychological Science*, *2*, 409-415.