bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465750; this version posted October 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity.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-ND 4.0 International license.

Internally Generated Time in the Rodent Hippocampus is Logarithmically Compressed

Rui Cao[1], John H. Bladon[2], Stephen J. Charczynski[1], Michael E. Hasselmo[1], and Marc W. Howard[1]

[1]Department of Psychological and Brain Sciences, Boston University
[2]Department of Psychological and Brain Sciences, Brandeis University

Abstract

The Weber-Fechner law proposes that our perceived sensory input increases with physical input on a logarithmic scale. Hippocampal "time cells" carry a record of recent experience by firing sequentially during a circumscribed period of time after a triggering stimulus. Different cells have "time fields" at different delays up to at least tens of seconds. Past studies suggest that time cells represent a compressed timeline by demonstrating that fewer time cells fire late in the delay and their time fields are wider. This paper asks whether the compression of time cells obeys the Weber-Fechner Law. Time cells were studied with a hierarchical Bayesian model that simultaneously accounts for the firing pattern at the trial level, cell level, and population level. This procedure allows separate estimates of the within-trial receptive field width and the across-trial variability. The analysis at the trial level suggests the time cells represent an internally coherent timeline as a group. Furthermore, even after isolating across-trial variability, time field width increases linearly with delay. Finally, we find that the time cell population is distributed evenly on a logarithmic time scale. Together, these findings provide strong quantitative evidence that the internal neural temporal representation is logarithmically compressed and obeys a neural instantiation of the Weber-Fechner Law.

Introduction

The Weber-Fechner law states that there is a logarithmic relationship between the perceived intensity of a sensory stimulus and the actual intensity (Fechner, 1860). The Weber-Fechner law provides a good description of a wide range of phenomena in psychology including perceptual phenomena (Moore, 2012), the non-verbal numerosity system (Feigenson, Dehaene, & Spelke, 2004; Gallistel & Gelman, 2000), and psychological studies of time perception (Rakitin et al., 1998; Gibbon, Church, & Meck, 1984). It has been proposed that

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465750; this version posted October 26, 2021. 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-ND 4.0 International license.

these psychological phenomena reflect logarithmically compressed neural representations of time, space, and number (Gallistel, 1989; Dehaene & Brannon, 2011). It has long been known that the representation of visual space in the retina and early visual cortical regions do not map visual space veridically. Rather, the representation of visual space has been argued to be logarithmically compressed as a function of distance from the fovea (Schwartz, 1980; Van Essen, Newsome, & Maunsell, 1984; Glezer, 1965). Moreover, neurons in monkey parietal cortex recorded during numerosity judgements have receptive fields organized as a function of the logarithm of numerosity rather than simply numerosity (Nieder & Miller, 2003). Similar findings are observed in human fMRI (Cantlon, Brannon, Carter, & Pelphrey, 2006), suggesting a common system across species (Nieder & Dehaene, 2009; Brannon, 2006). A recent paper from a cerebellar slice preparation argues that unipolar brush cells form a logarithmically-compressed set of temporal basis functions (Guo, Huson, Macosko, & Regehr, 2021). In hippocampus, it has been argued that a wide range of biological markers are expressed in log-scales (Buzsáki & Mizuseki, 2014). This paper asks the question whether hippocampal time cells form a logarithmic code for the time of past events, consistent with a Weber-Fechner representation of the past.

So-called "time cells" in the hippocampus fire sequentially during unfilled delay intervals (Pastalkova, Itskov, Amarasingham, & Buzsaki, 2008; MacDonald, Lepage, Eden, & Eichenbaum, 2011). Because the sequence is reliable across trials, one can reconstruct the time at which the delay began (e.g., Mau et al., 2018) and, in many cases, the identity of a stimulus that initiated the delay (Taxidis et al., 2020; Cruzado, Tiganj, Brincat, Miller, & Howard, 2020) such that at the population level hippocampal time cells construct a record of what happened when (Fig. 1a). It is well-established that time cells do not provide a veridical record of the past; rather the accuracy of the representation of time within the delay decreases with the passage of time (Kraus, Robinson, White, Eichenbaum, & Hasselmo, 2013; Mau et al., 2018; Taxidis et al., 2020). However, unlike the visual system, which clearly shows logarithmic compression, the quantitative form of the compression for time expressed in hippocampal time cells is not known. This question is central to our understanding of the function of the hippocampus in memory. Computational models of human episodic memory built from populations of time cells require that the compression of the time axis is precisely logarithmic (Howard & Eichenbaum, 2013; Howard, Shankar, Aue, & Criss, 2015). This form of compression is necessary to enable an account of the finding that the temporal properties of episodic memory are very similar over a very wide range of time scales, whether manipulated over seconds to minutes in the laboratory (Howard, Youker, & Venkatadass, 2008; Unsworth, 2008), over hours in experiments outside the laboratory (Mack, Cinel, Davies, Harding, & Ward, 2017), or studied using real-world events extending over much longer periods of time (Arzy, Adi-Japha, & Blanke, 2009; Uitvlugt & Healey, 2019). Logarithmic compression of memory makes sense of this observation for essentially the same reason that $\log(at) = \log a + \log t$. More precisely, logarithmic time compression means that the effect of rescaling time (taking $t \to at$) is to translate the memory to a new location rather than change the relative timing of events.

Logarithmic compression leads to two quantitative relationships (see Supplementary section S1, Fig. 1b, c) that have been empirically demonstrated in the visual system. First, the number of cells with receptive fields in a region of visual space goes down as the inverse of distance from the fovea (Daniel & Whitteridge, 1961; Hubel & Wiesel, 1974; Van Essen

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465750; this version posted October 26, 2021. 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-ND 4.0 International license.

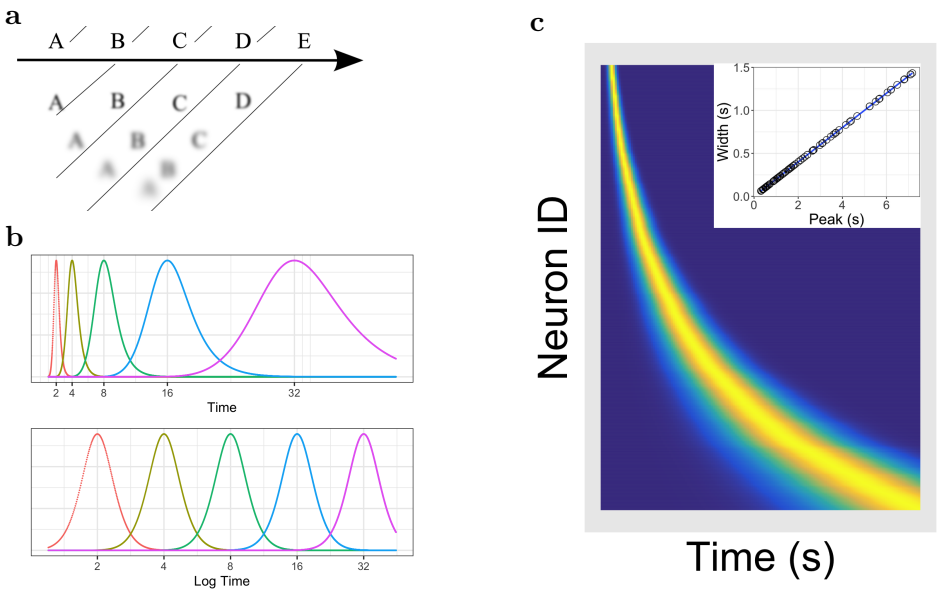

Figure 1. **Temporal representation and the Weber-Fechner law. a.** A compressed timeline of the past. The horizontal line A B C . . . represents objective reality as the experimenter presents a series of stimuli at regularly spaced intervals. At each moment, memory carries a subjective representation of what happened and when (diagonal line). The time of more recent events are represented more clearly while distant events are less distinguishable from each other, resulting in a compression of the temporal memory. **b.** A schematic illustration of a neural implementation of a logarithmically-compressed timeline. Each curve corresponds to a time field. When plotted as a function of log time (bottom), the receptive fields appear evenly distributed. When plotted as a function of linear time (top), the time-fields stretch out and become less numerous as time passes. **c.** Heat map of hypothetical time cells firing pattern if time is represented logarithmically. **insert** Time-field widths are linearly correlated with time-field peaks. Each row represents the firing pattern of a time cell, sorted by the peak of its time-fields. Dark blue indicates low activity while light yellow indicates high activity. Because time-fields follow logarithmic compression, the sorted peaks form a specific curved line where more time cells fire at the beginning of the delay. In the insert, the time-field widths are plotted as a function of their peaks. Logarithmic compression predicts a linear relationship.

et al., 1984). Second, the size of receptive fields grows linearly with distance from the fovea (Dumoulin & Wandell, 2008; Gattass, Gross, & Sandell, 1981). Although time cells are at least roughly consistent with these observations, a quantitative relationship has not been established. For instance, although it is well-established that receptive field widths go up with the peak time (e.g., Kraus et al., 2013) it has not been established that the relationship is linear and in any event prior work has estimated receptive fields by averaging over trials. This leaves open the possibility that the apparent growth in temporal receptive field width with the peak time is simply an averaging artifact (Fig. 5b).

The goal of the analyses in this paper is to rigorously test the hypothesis that time cells form a logarithmically-compressed representation of past time. We recorded from dorsal CA1 during a memory experiment in which rats had to remember the identity of an object across an eight-second delay (Fig. 4). This task is known to produce robust time cells,

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465750; this version posted October 26, 2021. 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-ND 4.0 International license.

which we identified using standard methods. We estimated the distribution of time field parameters at the level of individual trials, individual unit (across trials), and the population (across units) simultaneously using a hierarchical Bayesian model (Fig. 5). In hierarchical Bayesian modeling, each level of the hierarchy serves as a prior for parameter estimates at lower levels; parameters of the model at all levels are adjusted to maximize the likelihood of the entire observed dataset (Lee, 2011).

Results

Rats learned the memory task well and time cells, identified using previous methods (Tiganj, Cromer, Roy, Miller, & Howard, 2018), were identified in dorsal CA1. Because our goal was to isolate the effects of variability across trials from the within-trial time field, we restricted attention to time cells that fired on at least twenty trials. To mitigate edge effects in the power-law fitting procedure (Clauset, Shalizi, & Newman, 2009) we also required the peak time averaged over trials to be within 0.35 s and 7.6 s (see "Setting the bounds. . . " in Materials and Methods section for detail). A population of 131 robust time cells with time fields within the eight-second delay period were studied further.

To interrogate the neural representation of time from the trial-level to the population-level simultaneously, we applied a hierarchical Bayesian model to the selected time cells. We modeled time fields on each trial as a Gaussian. For each unit on trial i, the time field is centered at μ_i with standard deviation σ_w. We modeled μ_i as distributed according to a normal distribution across trials (eq. 3). The mean and standard deviation of the normal distribution were separately estimated for each time cell. Therefore, the mean M of this distribution estimates each time cell's average peak time. The standard deviation of this distribution, σ_t, estimates the variability of the location of the time field's location across trials. Logarithmic compression predicts that time field width goes up linearly with time field center. This will be evaluated by noting how the within-trial standard deviation σ_w varies across units as a function of M. Although the Bayesian analysis returns a posterior distribution for all of the parameters it estimates, to test hypotheses about relationships between parameters we will simply consider the mean of the posterior distributions as the model's estimate and use traditional statistics on those means.

Finally, the distribution of time field locations M across the CA1 population was assumed to follow a power-law distribution: $P(M = \tau) \propto \tau^{-\alpha}$. The Bayesian model returns the posterior distribution of α given the data. That means that this approach is capable of observing logarithmic compression—a tight posterior distribution around $\alpha = 1$—but also other hypotheses as well. For instance, $\alpha = 0$ corresponds to a uniform distribution of time field locations (Fig. 1e); every other positive real number is also possible.

The hierarchical Bayesian model provided a qualitatively satisfactory description of time-specific spiking across trials and across neurons. Example model fits of individual cells across trials can be found in Figure 2a (see also Supplementary Figure S2). For ease of exposition, the results are presented in two sections. The first section ("The representation of time varies . . . ") focuses on variability in time fields across trials. The second section ("Hippocampal time cells form a . . . ") focuses on the properties at the population level and provides direct tests of the hypothesis that the hippocampal time code is logarithmically compressed. It should be emphasized that although these results are presented sequentially,

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465750; this version posted October 26, 2021. 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-ND 4.0 International license.

the parameters of the model were estimated using data from both levels of description simultaneously.

The representation of time varies from trial to trial and is internally consistent across units

Our model fits allowed a close examination of the temporal representation at the trial-level, which was ignored in previous studies of time cells. It is apparent that firing patterns change from trial to trial. The model captures the trial variability both by providing trial level time field estimation, μ_i on each trial i, and by quantifying the across trial variability for each cell *via* σ_t.

There are three novel findings reported in this section. First, there was robust variability in time field locations across trials. This is consistent with several hypotheses. It is possible that time cells are simply noisy. A more interesting hypothesis is that the animal's internal sense of time, although correlated with clock time on average, varies from moment to moment and from trial to trial. Second, the locations of time field centers for simultaneously-recorded time cells recorded from different tetrodes were correlated across trials as if the population cooperated to construct an internal sense of time. Finally, we found that across-trial variability changed systematically as a function of time within the delay. This means that compression observed in previous studies of time cells is at least partially attributable to a trial averaging artifact. As we will see below, there is also robust evidence for compression that is not attributable to a trial averaging effect.

Time field location varies from trial to trial. Our time field model revealed a high degree of variability in the estimated time field location from trial to trial. Figure 3a plots representative single units with trial-by-trial model fits for cells that fired at different times during the delay interval. As a control to ensure that the time field location varied meaningfully across trials we compared this model to an alternative model where the between-trial variability was captured by varying the width of the time field, rather than the peak time. A large majority of cells (91/131, $\chi^2(1) = 19.1$, $p < .001$) were better fit by the model in which time field location varied across trials (see Supplementary section S2, S1). We conclude that estimates of time field location across trials are a meaningful estimate of variability across trials.

An internally coherent temporal representation in hippocampus. The variability in the location of time fields from trial to trial means that, to the extent time cells are measuring time, this internal estimate does not precisely map onto the objective time measured by the experimenter using a clock. On each trial, one could ask whether such variability is a product of stochastic noise, or if it is the product of a meaningful fluctuation in the internal state. If variability across trials is the result of a stochastic computation at the level of the neuron, trial-level estimated time fields should not be correlated across simultaneously-recorded cells. By contrast, if the population cooperates in constructing a coherent estimate of time, albeit one that is not perfectly correlated with the readings on the experimenter's clock, this would predict that simultaneously recorded units vary their time fields across trials in coordination with each other. To address this question, we investigated trial-by-trial correlations of simultaneously recorded pairs. To avoid spike sorting errors we omitted pairs of time cells that were recorded from the same tetrode.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465750; this version posted October 26, 2021. 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-ND 4.0 International license.

Many simultaneously recorded unit pairs appeared to shift together from trial to trial. Figure 2b plots the rasters of a pair of simultaneously recorded time cells across trials as an example. This pair of time cells (and all pairs subsequently considered) were not recorded from the same tetrode. Visual inspection of the raster suggests that the time fields move together across trials. For instance on trial 19 both time cells fired relatively later than usual. In the insert, we plotted the time field locations μ_i estimated by the model for both members of the pair of time cells. Each point in the insert is one trial. It is clear that the model estimates of μ_i for these two time cells are correlated across trials ($r(24) = .83$, $p < .001$).

To test whether time field locations shifted from trial to trial in concert with one another, we calculated the Pearson correlation between μ_i for all simultaneously recorded time cell pairs that were not from the same tetrode. Of the 185 simultaneously recorded pairs, three were excluded for low numbers of observations (< 10 simultaneously recorded trials). The average number of shared trials between the remaining 182 pairs was 32.4. The Pearson correlation between those pairs was positive for a majority of the pairs, ranging between [-.51, .98] with the median at .108 (brown curve in Figure 2c). We applied one-sample t-test (although the independent sampling assumption is violated because many cells appeared in more than one pairs) and found the mean of correlation distribution is significantly above zero: $t(181) = 6.09$, $p < 0.001$. As a negative control, we also performed the same correlation test between randomly paired cells from different sessions as if they were recorded simultaneously. When compared via Kolmogorov-Smirnov test, the simultaneous pairs exhibited a higher correlation in shifts than the control pairs ($p < .001$, see Figure 2c). Because the randomly shuffled pairs still preserved the order of trials, the fact that correlations between simultaneously-recorded pairs are significantly higher suggests that the correlation is unlikely due to other variables confounded with passage of time within the session (e.g., build-up of fatigue or satiety). These results suggest that the fluctuations in the representation of time across trials reflect changes in a meaningful neural process.

Estimated across-trial variability changes with delay. The hierarchical Bayesian model enables a separate estimate of within-trial variability—σ_w, the width of the time field within trial—and across-trial variability, σ_t. Discussion of within-trial variability is postponed to the next subsection. For each time cell, σ_t quantifies how much the time field shifts across trials. Although the hypothesis that the temporal record of the past is logarithmically-compressed at each moment is agnostic about the question of how time is represented across trials, it is nonetheless important to quantify across-trial variability in the attempt to isolate the effect of within-trial variability on trial-averaged time fields. The model identified substantial across-trial variability. The left panel of Figure 2e plots the distribution of σ_t across time cells. The median value of σ_t was .69 s. Thirty-four percent of time cells showed an across-trial standard deviation of more than 1 s. The model's estimate of across-trial variability was larger than its estimate of the width of the time field for a substantial majority of time cells (93/131, $\chi^2(1) = 22.3$, $p < .001$).

There was also a systematic change in across-trial variability as a function of the peak location (Fig. 2e). There are a number of uninteresting factors that could affect this quantitative relationship. For instance, for a M that is near the edge of the delay interval, it would be more difficult to estimate a large value of σ_t because the distribution would be cut off by the edge of the delay. With this caveat in mind, we applied regression analysis

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465750; this version posted October 26, 2021. 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-ND 4.0 International license.

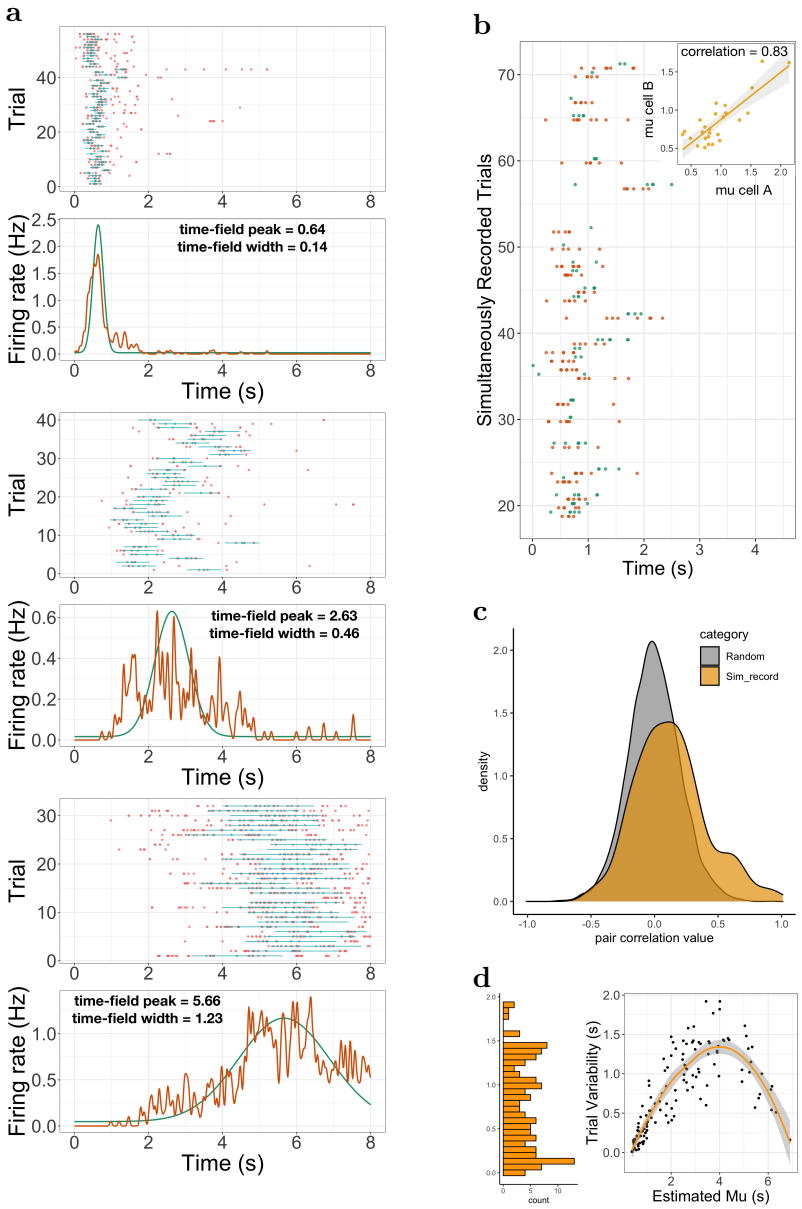

Figure 2. **Internal timeline varies across trials but is correlated within-trial.** **a.** Three representative time cells with different temporal receptive fields across the delay. Rasters show estimated time fields for each trial as a blue line. In the bottom plots, the red line represents the peristimulus-time histogram averaged over trials. The green line indicates the estimated time field with within-trial time field width (σ_w, shown above). The across trial location variability σ_t is estimated separately in the model. **b.** Raster for a pair of simultaneously-recorded time cells from different tetrodes. Different time cells are shown in different colors. Inset: time field peaks estimated by the model for each time cell on each trial. The peak times of this pair of time cells were reliably correlated across trials. **c.** Distribution of Pearson correlations for all simultaneously-recorded time cell pairs from different tetrodes (brown distribution) and a shuffled distribution (grey distribution). **d.** Estimated trial variability σ_t as a function of the peak firing time M. The left histogram plots the distribution of the trial variability. The line shows the results of a linear regression including constant, linear and quadratic terms (see text for details).

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465750; this version posted October 26, 2021. 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-ND 4.0 International license.

of trial variability (σ_t) as a function of the time field peak (M), including an intercept and linear and quadratic terms. All three coefficients were significant (intercept, Mean ± SE: $-.33 \pm .06$, $p < .001$; slope: $.84 \pm .05$, $p < .001$; quadratic term: $-.11 \pm .01$, $p < .001$; $R^2_{adj} = .73$). We conclude from this that variability in the location of the time field across trials varies systematically with the average location M.

Because variability in the location of a time field across trials is substantial and varies systematically with time within the delay, quantitative estimates of time field width as a function of time field center should control for this trial variability. Fortunately, the hierarchical Bayesian analysis provides separate estimates of within- and between-trial variability. We present results for time field width isolated from trial averaging effects in the next subsection.

Hippocampal time cells form a logarithmically compressed timeline of the past

In the last subsection we found that the animals' internal representation of time is internally consistent across the population within a trial but varies substantially across trials. After accounting for the across-trial variability, we are now able to study the form of temporal compression of the time code within a trial. If populations of time cells represent a compressed timeline, we should observe fewer units with larger values of M and the time field width σ_w should increase with M. This would mean that the temporal resolution carried by the population is reduced as the event fades into the past. If the population is compressed logarithmically, this makes two quantitative predictions. First, time field width σ_w should increase linearly with the delay M. Second, time field peaks M should be distributed according to a power-law with an exponent of 1. Section S1 of the supplementary materials provides an elementary explanation of why logarithmic compression leads to these properties.

Time field width increases linearly with the delay. Figure 3b plots within-trial time field width σ_w as a function of the peak location of the time field M for each time cell. Examination of this plot suggests that time field width increases with delay, even after taking the across-trial variability into consideration. To quantify this relationship, we fit a regression analysis of time field width σ_w as a function of time field peaks M time field width including constant, linear, and quadratic terms. Only the the linear term reached statistical significance ($.14 \pm .05$, $p < .005$, $R^2_{adj} = .21$). The intercept ($.11 \pm .06$, $p < .07$) and quadratic ($-.007 \pm .008$, $p > .4$) terms were small and neither reached traditional levels of significance. The reliable linear term indicates that time field width increases with the peak time even after controlling for trial-averaging effects.

Logarithmic compression predicts a precisely linear relationship between within-trial time field width σ_w and peak time M. While there was no evidence for a quadratic term, the intercept did approach significance. There is no question, however, that the curve cannot literally go to the point (0,0). If $M = 0$ it would require that hippocampal time cells respond instantaneously to the flag indicating the beginning of the delay, which is not possible. Moreover, it would be impossible to estimate a σ_w approaching zero—it makes no sense to estimate a standard deviation without more than one spike. As it is, even if the intercept had reached significance at the upper limit of the confidence interval we observed (on the order of a few hundred ms), we would still conclude that the intercept is small in magnitude relative to the range of delay times exhibited by time cells in this experiment.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465750; this version posted October 26, 2021. 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-ND 4.0 International license.

Time cell peaks are distributed according to log compression. Figure 3a shows the firing rate, averaged over trials, for each of the time cells sorted by their estimated peak time M. Consistent with many prior studies of time cells, the central ridge appears curved. The curvature is a sign of a non-uniform distribution of time field peaks M—if M was uniformly distributed across time, one would expect to see a straight line when plotted in this way. As observed in prior studies, the shape of the curvature appears consistent with logarithmic compression—the dashed white line on Figure 3a shows the curve that would be expected if the time fields were distributed uniformly over log time. The hierarchical Bayesian model enables us to quantify this relationship more precisely.

The hierarchical model models the distribution of time field peaks across the population as $M \sim M^{-\alpha}$ (see Eq. 4 and methods for more details). When $\alpha = 0$, the peak of time cells should be distributed evenly throughout the delay, showing no temporal compression. For $\alpha > 0$ the representation is compressed. Crucially, $\alpha = 1.0$ indicates that the compression is exactly logarithmic.

Figure 3c plots the posterior distribution of α. This distribution is centered is around 1 (mean = .92 with 95 % interval [.67, 1.27]). This posterior conclusively argues against uniformly distributed time cells as it excludes zero by approximately five standard deviations. It also argues strongly against other "simple" values for α that one might have considered reasonable *a priori*, including 1/2, 3/2 and 2. To informally assess how well the distribution of M approaches $\alpha = 1$, Figure 3d plots the cumulative distribution function (CDF) of time cell peaks M on a log axis. If α were 1.0, the CDF would appear as a straight line when plotted in this way. It should be noted that the estimates of M for different time cells are not independent of one another in the hierarchical Bayesian analysis. However, one also finds similar results from traditional estimate of time fields (Figure S3d), at least considering fields that peak after a few hundred ms.

Discussion

The present study analyzed a group of hippocampal CA1 time cells with a hierarchical Bayesian model that describes the firing patterns simultaneously at the individual trial level, the individual cell level, and the population level. There were three important observations from the model parameter estimates. First, by analyzing the model parameters at the individual trial level, we found that the time fields of simultaneously recorded time cells correlated with each other from trial to trial. This suggests an internally coherent time representation within each trial. Second, analysis at the individual cell level revealed that time field width increased linearly with delay, even after taking the across-trial variability into consideration. Third, estimating the power-law exponent that describes the population of time field peaks resulted in a posterior distribution close to one; an exponent of 1.0 means the compression of the internal timeline is logarithmic. These results provided quantitative evidence that the temporal representation in hippocampus was internally coherent across trials and logarithmically compressed within a trial.

Potential causes of variability across trials

There are a number of external and internal factors that may have contributed to the variability in the neural representation of time across trials. First, at least some of the

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465750; this version posted October 26, 2021. 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-ND 4.0 International license.

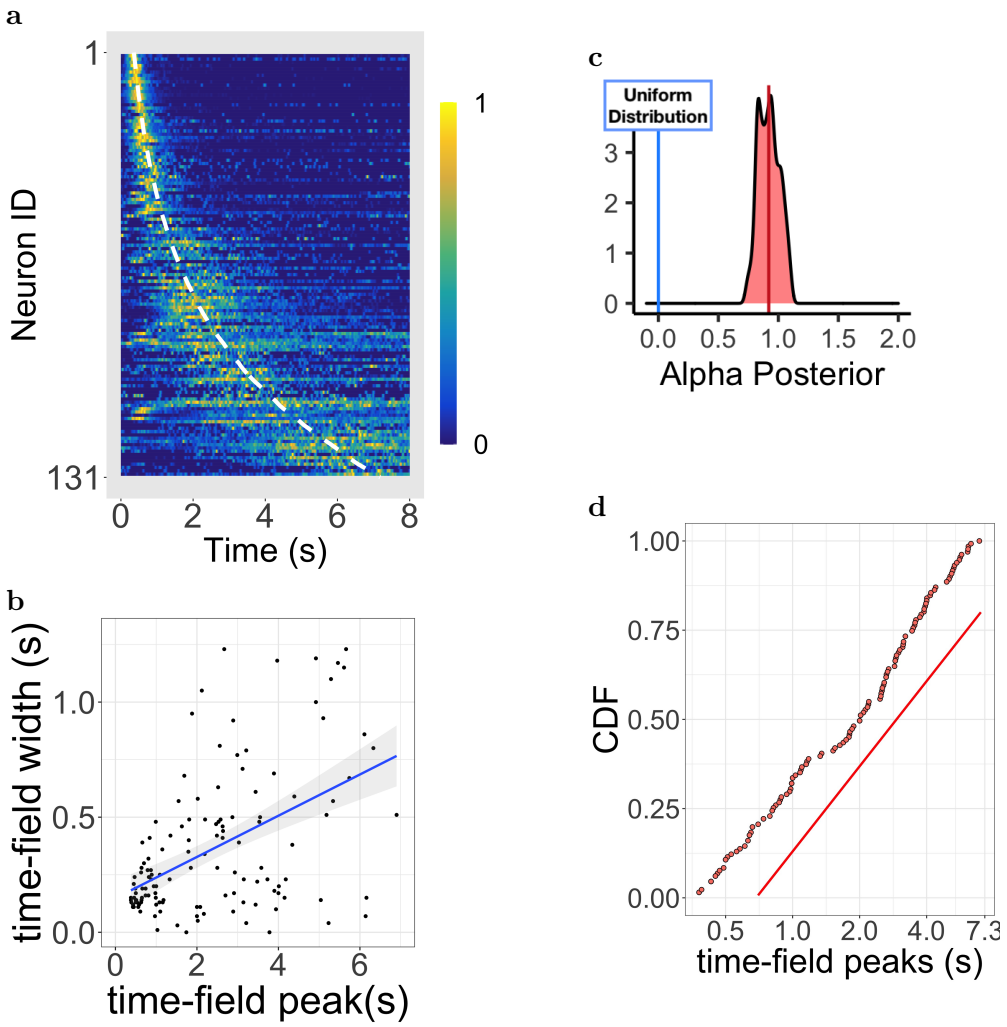

Figure 3. **Log-compressed timeline a.** Normalized firing rate of time cells sorted by peak time. Yellow indicates high activity and blue indicates low activity. The dashed white line shows the peak times one would obtain if time cells uniformly sampled log time. Equivalently, the dashed white line shows the peak times that would result if the probability of observing a peak time M went down like M^{-1} corresponding to $\alpha = 1.0$. **b.** The width of the within-trial time field as a function of the peak firing time. The line shows the results of a linear regression including constant, linear and quadratic terms. Neither the intercept nor the quadratic term reached significance (see text for details). **c.** The posterior distribution of the power-law exponent parameter α. The blue line on the left marks zero, corresponding to a uniform distribution. The red line marks the mean (.93) of the posterior. **d.** Cumulative distribution function (CDF) of time field peaks M plotted on a log scale. The straight red line is included to facilitate visual comparison.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465750; this version posted October 26, 2021. 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-ND 4.0 International license.

variability could be attributable to variability in the process or processes that initiate the sequence of time cells. Second, at least some of the variability could be attributable to the rate at which the sequence proceeds.

Possible sources of variability in the initiation of time cell sequences. Measuring time implies expressing the relationship between the present and some remote time points. For instance, in the afternoon, a traditional clock (i.e., one with three hands and twelve numbers) records the hours, minutes and seconds since noon. In this paper, we defined time during a trial as the time that elapsed since the animal broke an IR beam to enter the treadmill. But the population of time cells are not necessarily triggered by that event, nor is it necessary that all time cells are triggered by a single event. Subtle changes in the way the animal broke the beam may have generated variability in the rats' physical and/or perceived position the moment the treadmill was triggered. Perhaps some time cells are triggered by the impact of the paw on the treadmill, which follows the beam break, or the step towards the treadmill, which precedes it. There are a number of other possible events clustered around the moment of the beambreak that could trigger some or all of the time cells.

Possible sources of variability in the rate of time cell sequences. Given a starting point that varies from trial to trial, it is also possible that the relationship between time measured by a clock and time measured by a population of time cells is not constant from moment to moment. Many factors affect subjective duration estimates in humans (Eagleman, 2008) and a number of neurobiological factors including dopamine release (Soares, Atallah, & Paton, 2016) and theta frequency (MacDonald, Carrow, Place, & Eichenbaum, 2013) could affect the rate at which neural estimates of duration unfold. We are unable to distinguish how these factors may have impacted this experiment. However, regardless of the source of variability in the flow of time, we can make predictions about the shape of the function relating across trial variability in time field location to the mean time field location across trials.

Suppose that within a trial the current estimate of internal time updates at each moment as $\Delta + \eta$ where Δ is a positive constant from moment-to-moment and η is a symmetric random variable with zero mean and variance σ^2. The randomness due to η causes the rate at which time flows to vary within a trial. Under these circumstances, by the central limit theorem we would expect after T timesteps the internal interval to be distributed like $\Delta T + \mathcal{N}\left(0, \sigma\sqrt{T}\right)$. That is the mean goes up linearly with T, but the standard deviation across trials goes up with $\sqrt{T}$. Indeed, we found a sublinear relationship between σ_t and M, at least roughly consistent with this argument. Regardless of the source of the variability in the rate at which time flows from moment to moment—i.e., what unmeasured variables affect η—we should nonetheless find that variability in μ_i across trials should be a sublinear function of M.

Compressed representations of time, space and other variables in hippocampus and related regions

This paper argues that within a trial the internal timeline expressed by the population of time cells is logarithmically compressed. Although the form of the distributions were not assessed as carefully, many other studies of time in the hippocampus and related regions

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465750; this version posted October 26, 2021. 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-ND 4.0 International license.

show a broad spectra of timescales and evidence of compression. That is, many prior studies show a graded heterogeneity of time scales across units and time scales close to zero are overrepresented relative to time scales representing times further in the past. The arguments for the optimality of logarithmic compression are not only relevant for time but apply equally well to any one-dimensional variable with a natural zero. Distance and natural number are obvious examples of such variables. There is ample evidence suggesting that representations of space in the hippocampus and related regions could be logarithmically compressed as well.

Diverse timescales across brain regions. Many previous papers have shown compression of hippocampal time cells in rodent hippocampus consistent with log compression (e.g., Kraus et al., 2013; Mau et al., 2018; Taxidis et al., 2020). In addition, a qualitatively similar compression has been reported for sequentially-activated cells in rodent striatum (Mello, Soares, & Paton, 2015), rodent mPFC (Tiganj, Shankar, & Howard, 2017), monkey caudate nucleus and DLPFC (Jin, Fujii, & Graybiel, 2009), monkey lPFC (Tiganj et al., 2018; Cruzado et al., 2020) and monkey hippocampus (Cruzado et al., 2020) during a broad variety of behavioral tasks. It has recently been reported that unipolar brush cells in the cerebellum form a logarithmically-compressed temporal basis set in a slice preparation (Guo et al., 2021). It is theoretically important to systematically evaluate whether these regions show the same quantitative distributions; this would require careful experimentation. Because the experimental challenges of estimating time are greatest for times near zero, it may be preferable to conduct this experiment over time scales greater than a few seconds; it has been argued that time cells fire sequentially over several minutes (Shikano, Ikegaya, & Sasaki, 2021; Liu et al., 2021).

In addition to sequentially-activated time cells, cells in the entorhinal cortex can be used to decode time since a triggering event. Rather than fire sequentially with graded heterogeneity of peak times, cells in EC, which have been referred to as temporal context cells, appear to change their firing rate gradually following an exponential function as the triggering event recedes into the past. Recording from rodent LEC in a spatial exploration task Tsao et al. (2018) observed cells that gradually changed their firing rate with a wide range of time constants spanning up to several minutes. The graded heterogeneity in time constants of decay (or ramp) allow decoding of time since an event over a wide range of time scales. Bright et al. (2020) observed similar results in monkey EC during a free viewing task with cells triggered by onset of an image and decaying with time constants ranging from a few hundred ms up to several s. Entorhinal neurons that are not typically understood to be coding for time also show a graded heterogeneity of time constants. Dannenberg, Kelley, Hoyland, Monaghan, and Hasselmo (2019) recorded from speed cells in rodent MEC. Speed during exploration changes continuously; that study asked over what time scale filtering the speed signal provided the best account of each speed cell's time course of firing. Rather than finding a single value, different speed cells appeared to filter the speed signal with a wide range of time constants. This distribution overrepresented times near zero and showed a long tail, much like the distribution of time constants in the EC studies described above and not unlike the logarithmic distribution of time cell peaks observed in this study. Although there are many possible biological mechanisms for graded heterogeneity of time constants, slowly-inactivating currents in EC slice experiments (Egorov, Hamam, Fransén, Hasselmo, & Alonso, 2002) is an attractive candidate for the graded heterogeneity of time constants

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465750; this version posted October 26, 2021. 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-ND 4.0 International license.

in EC (Tiganj, Hasselmo, & Howard, 2015; Liu, Tiganj, Hasselmo, & Howard, 2019).

Compressed spatial representations in hippocampus and related regions. In this study, we saw that time fields are more numerous around the beginning of the delay (e.g., Fig. 3). This result replicates many previous studies (e.g., Kraus et al., 2013; Mau et al., 2018; Taxidis et al., 2020). There is also evidence suggesting that place fields cluster relative to landmarks in the environment, such as the beginning of a path (Bjerknes, Dagslott, Moser, & Moser, 2018; Sheehan, Charczynski, Fordyce, Hasselmo, & Howard, 2021) and edges of an open field environment (Harland, Contreras, Souder, & Fellous, 2021). In order to make a clear comparison between temporal representations and spatial representations, it is essential to establish the "zero" of the spatial coordinate system. That is, populations of time cells appear to code for the time between a particular event and the present. However, when an animal is in any specific location in a spatial environment, many different landmarks are at different distances and headings. In order to directly compare time to space, it is essential to establish experimental control over the place code by moving a landmark in the environment and observing that the place cells move with the landmark. Sheehan et al. (2021), following a method introduced by Gothard, Skaggs, Moore, and McNaughton (1996); Gothard, Hoffman, Battaglia, and McNaughton (2001), systematically manipulated the location of a reward box along a linear track. As observed by Gothard et al. (1996, 2001), many hippocampal place fields moved with the box, coding for distance traveled since leaving the box. Consistent with the results from time cell experiments, place fields tiled the distance around the box smoothly. Moreover, place fields were more numerous and more narrow closer to the landmark controlling their firing than further away from the landmark (Sheehan et al., 2021). It is not currently known whether the form of the compression of the place code is also logarithmic, but the results are at least qualitatively similar to those observed for time cells.

Work on grid cells has shown quantitative evidence of a form of compression of space that is closely related to the logarithmic compression observed for time in this study. It has long been appreciated that the spatial frequency of grid cells falls into discrete groups (Stensola et al., 2012). Interestingly, the relative ratio between the spatial frequency of adjacent clusters of grid cells is constant, as predicted by a normative account of spatial coding (Wei, Prentice, & Balasubramanian, 2015). Logarithmic compression also implies that the ratio between adjacent time scales are constant (see Eq. S2). That is, both spatial frequency of grid cells and temporal receptive fields for time cells form a geometric series. The difference is the number of cells that share each value. Whereas time cells appear to smoothly cover the log time axis, many grid cells cluster at a few evenly-spaced locations on the log spatial scale axis. Whereas the ratio between adjacent time scales is close to one, the ratio of the spatial frequency of one cluster of grid cells to the next is much larger. It is an interesting theoretical problem to understand why the brain has chosen distinct but closely related coding schemes for time (and perhaps also space) in the hippocampus and spatial frequency in the grid cell system.

Implications of logarithmically-compressed time for cognitive neuroscience

Human episodic memory—the vivid recollection of a specific event from one's life—has long been thought to depend on recovery of a prior state of spatiotemporal context (Tulving & Madigan, 1970). Computational models of human list learning experiments built on

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465750; this version posted October 26, 2021. 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-ND 4.0 International license.

this assumption (Sederberg, Howard, & Kahana, 2008; Polyn, Norman, & Kahana, 2009; Talmi, Lohnas, & Daw, 2019; Logan, 2021) provide a natural account of the contiguity effect—the finding that memory for a particular item from a list spontaneously brings to mind other stimuli presented at nearby points in space and time (Healey & Kahana, 2013; Kahana, 2012). The hippocampus has been proposed as a brain region that both maintains the current state of temporal context and is essential for the recovery of temporal context hypothesized to support episodic memory (for a review Eichenbaum, 2017). It has been proposed that a population of logarithmically-compressed time cells could serve the role of temporal context in computational models of human episodic memory (Howard & Eichenbaum, 2013; Howard et al., 2015). The empirical observation that time cells form a logarithmically-compressed representation of the recent past has dramatic implications for computational models of human episodic memory.

Retrieved temporal context in human MTL. According to many computational models of episodic memory in humans, episodic memory recall is mediated by a distributed pattern of temporal context that changes gradually over time. This property is essential for its role in accounting for temporal effects in human list learning studies. In parallel with animal work, a number of human neuroimaging studies show evidence for a gradually-changing representation of the recent past in human hippocampus and related regions (Tubridy & Davachi, 2011; Cohen, Ross, & Stern, 2018; Hsieh, Gruber, Jenkins, & Ranganath, 2014; Deuker, Bellmund, Schröder, & Doeller, 2016; Schapiro, Turk-Browne, Norman, & Botvinick, 2016; Bellmund, Deuker, & Doeller, 2019).

These same computational models of episodic memory argue further that the behavioral contiguity effect results from recovery of a previous state of temporal context—a neural jump back in time. If the behavioral contiguity effect results from recovery of a previous state of a gradually-changing temporal context, then it should be possible to observe a neural contiguity effect as well. The pattern of activity when participants successfully recall a studied item should resemble the pattern of activity during study of the recalled item, but crucially also resemble the pattern during study of neighbors of the recalled item. Indeed, neural similarity decreases as function of the temporal distance between the studied items and its neighbors, showing a neural contiguity effect (Folkerts, Rutishauser, & Howard, 2018; Manning, Polyn, Litt, Baltuch, & Kahana, 2011; Yaffe et al., 2014; Nielson, Smith, Sreekumar, Dennis, & Sederberg, 2015). In a particularly dramatic example, Nielson and colleagues (2015) asked participants to wear smartphones around their necks as they went about their daily lives over a month-long experiment. The smartphone took pictures of their environment at regular intervals with corresponding time stamps and GPS information for a month. The images from participants' own life were later presented to them in the scanner with the instruction to "try to mentally relive your experience." The pattern similarity of hippocampal activition between pairs of remembered pictures correlated the the objective temporal and spatial distances between the actual experiences. Critically, pattern similarity in hippocampal activation for retrieved memories correlated with temporal and spatial distances over temporal scales of weeks—much longer than would typically be observed in the laboratory.

Log compression of temporal context leads to scale-invariance of contiguity effect. If temporal context is constructed from time cells, and the contiguity effect results from recovered temporal context, then the logarithmic compression observed in this paper

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465750; this version posted October 26, 2021. 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-ND 4.0 International license.

makes sense of an important and otherwise puzzling property of human episodic memory: The contiguity effect in episodic memory is observed across a very wide range of time scales. For instance, in a delayed free recall experiment in which the items are presented once per second, a robust contiguity effect is observed (Kahana, Howard, & Polyn, 2008). However, if one inserts a distractor-filled delay of, say, sixteen seconds between each word, presumably clearing the contents of short-term memory between each list item, a contiguity effect is still observed (Howard & Kahana, 1999). If, at the end of an experimental session lasting tens of minutes, one asks the participant to recall all the words from all the lists, there is a contiguity effect *across lists*, despite the fact that the words from different lists are separated by perhaps a minute (Unsworth, 2008; Howard et al., 2008). It is a computational challenge to understand how the contiguity effect can persist across all of these time scales.

If temporal context is constructed from logarithmically-compressed time cells, the logarithmic compression provides a natural explanation for this phenomenon. This paper argues that time cells in the rodent hippocampus record past events as a function of $\log t$. Recall that $\log(at) = \log a + \log t$. That is, multiplying the time axis by some constant scaling factor a shifts a logarithmic scale by $\log a$. For a logarithmically-compressed timeline, rescaling time—zooming out from the scale of one second to ten seconds to one hundred seconds—has the same effect as simply shifting to a new set of time cells. Any associative network built from a set of logarithmically-compressed time cells will respond to rescaling time in the same way. If one can devise a method to pattern complete activation of a set of time cells driven by stimuli presented at one time scale, the same network will also have a similar ability to pattern complete when the sequence of stimuli are rescaled in time. Notably, this same property of logarithmically-compression also leads naturally to time-scale invariant perception when integrated into deep neural networks (Jacques, Tiganj, Sarkar, Howard, & Sederberg, 2021), suggesting that logarithmic compression is a general principle of temporal coding in the brain (Guo et al., 2021).

Although logarithmic compression of time cells provides a solution to part of the computational puzzle posed by the behavioral contiguity effect, there are still significant challenges. For instance, the contiguity effect extends over time scales much longer than have currently been observed in time cell experiments (which peak at perhaps a few minutes Shikano et al., 2021; Liu et al., 2021). The contiguity effect can be observed when stimuli are presented separated by an hour (Mack et al., 2017). Although it is difficult to do a controlled experiment, Uitvlugt and Healey (2019) observed contiguity effects that extended over weeks and months when asking participants to recall news events. This recalls the findings of Nielson et al. (2015), who observed neural contiguity effects extending over similar time scales for autobiographical memories. Arzy et al. (2009) asked subjects to place real world events on a timeline centered at the present. They found the accuracy for the time of past events decreased linearly with log-scaled time, suggesting a logarithmically-compressed temporal representation over time scales up to decades (see also Peer, Salomon, Goldberg, Blanke, & Arzy, 2015). Although mathematically a logarithmic scale can represent events arbitrarily far in the past, it seems unlikely that the same neurophysiological processes that lead to sequentially-activated time cells over scales from about a hundred ms to perhaps tens of minutes could extend over such long time scales. Perhaps representational drift, which can be observed for very long time scales (Ziv et al., 2013; Rubin, Geva, Sheintuch, & Ziv, 2015; Cai et al., 2016; Mau et al., 2018; Mau, Hasselmo, & Cai, 2020) is responsible for

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465750; this version posted October 26, 2021. 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-ND 4.0 International license.

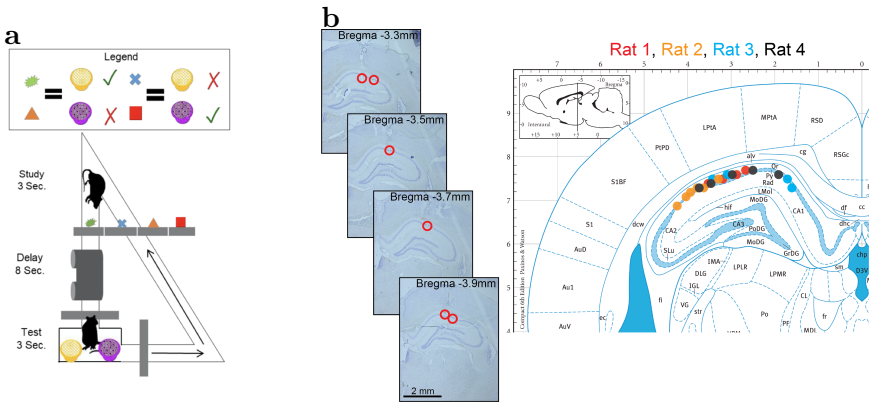

Figure 4. **Data acquisition a** Illustration of the behavioral task. **b** Recording sites. The left panel plots example histology sections for each animal. the red circle indicates the final recording location. The right panel plots the estimated tetrode locations for each rat (coded by different color dots) mapped onto the Paxinos and Watson atlas (1986).

the contiguity effect in episodic memory over autobiographical time. If representational drift codes for events over log time, as we observed for hippocampal time cells here, then it should be possible to trigger very slow sequences with specific environmental events (Liu et al., 2021).

Materials and Methods

Experiment and Electrophysiology

Subjects. Subjects were four male Long-Evans rats (Charles River) weighing between 350 and 450 g and between the ages of 6 months to 1.5 years for the duration of the experiment. Animals were single housed and maintained on a 12-hour light-dark cycle (lights on at 8:00 AM) for the duration of the experiment. All behavioral training and experimentation were performed during the light phase. Animals were given ad-libitum water and maintained at a minimum of 85% of their ad libitum feeding body weight during all behavioral training and testing. All procedures were conducted in accordance with the requirements set by the National Institutes of Health, and were approved by the Boston University Institutional Animal Care and Use Committee (BU IACUC).

Behavioral apparatus. The behavioral apparatus consisted of a custom-built 355 cm long by 7.5 cm wide circular track with an integrated treadmill using commercially available parts (Columbus Instruments). The treadmill had walls funneling into a small exit to ensure the animals' head position was fixed for the duration of the treadmill run. At the end of the treadmill existed a 20×20 cm square platform onto which the test objects were placed. The track was elevated 95 cm above the floor, positioned at least 20 cm away from all walls but in a position in which various distal visual cues were available to the rat. The maze contained two automatic doors, one at the front of the treadmill and one on the return arm that was controlled by infrared beam breaks (Arduino microcontroller, hardware from Adafruit industries).

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465750; this version posted October 26, 2021. 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-ND 4.0 International license.

Training procedure. Rats were trained in a similar manner to previous experiments (Robinson et al., 2017). Briefly, rats were initially trained to run in one direction around the maze for froot-loop reward. Once rats reliably looped, they were shaped to retrieve the reward hidden in a flower pot each lap (roughly 4 days). Then, rats were trained to sample one of two test objects before running through the treadmill platform and then choosing the matching one of two terra cotta pots discriminable by scent and digging media (Keene et al., 2016). The terra-cotta pots were always placed side-by-side on the platform in a pseudorandomized position. A choice was determined to be made once the animal disturbed the surface of the media at which point the opposite pot was immediately removed at this and all subsequent stages of behavior. One crushed froot loop was added to each pot at the beginning of each day, and every 7-9 trials neither pot contained a food reward, and the reward was given after 1 second of digging to prevent the possibility of reward scent guiding behavior. This shaping was performed by initially running each object-pot pair in large blocks of trials (>10 consecutive trials) and progressively weaning down to random presentation over the course of 2 weeks. Once animals were performing at >80% discrimination, two new objects were added one at a time to the study set using the same progressively shrinking block design. Once animals were able to discriminate between the 4 objects to retrieve reward from the correct terra-cotta pot of the two with >90% accuracy, a running delay was imposed. Initially, animals were trained to wait for a fraction of a second while the treadmill was nudged forward between the study and test phases. This treadmill delay was progressively increased until performance fell to between 70 and 80%, typically 8 seconds. Typically, this training schedule took roughly 2 months to train animals on the first pair of study objects, and then another 1.5 months for rats to perform with 4 study objects and with a sufficiently long treadmill delay.

Surgery. Anesthesia was induced via inhalation of 5% isofluorane (Webster Veterinary Supply) in oxygen, and then a stable plane was maintained at 1.5-3% throughout the entire surgery. Immediately following the induction, animals were injected with the analgesic Buprenex (Buprenorphine hydrochloride, 0.03 mg/kg i.m.; Reckitt Benckiser Healthcare), and the antibiotic cefazolin (330 mg/ml i.m.; West-Ward Pharmaceuticals. The rat was then fixed to a stereotaxic frame (Kopf). Craniotomies then were made above the dorsal hippocampus (dHPC) (AP-4.1, ML 3.0 mm), and the rostral lateral entorhinal cortex (LEC) (AP-6.8, ML 4.5). Six to eight sterile stainless steel screws were then fixed to the skull, and the remaining skull surface was covered with Metabond (Parkell, Inc). A hyperdrive consisting of 24 independently moveable tetrodes was then lowered into the craniotomy, and fixed to the skull using dental cement. Two ground screws were inserted above the cerebellum, and soldered to the ground leads on the electrode interface board during surgery. Each tetrode was composed of four 12 uM RO 800 wires (Sandvik Kanthal HP Reid Precision Fine Tetrode Wire; Sandvik). Tetrodes were plated with non-cyanide gold solution (Neuralynx) via electrolysis using a Nano-Z (White Matter LLC) in order to reduce impedance to between 180 and 220 kOhms. At the conclusion of surgery tetrodes were lowered 1 mm into dorsal cortex. Animals were given Buprenex and Cefazolin twice a day as needed for up to three days post-surgery.

Data acquisition. All electrophysiological recordings were performed using a 96 channel multichannel Acquisition Processor (MAP) recording system (Plexon). Each spike channel was amplified 1000x, then between 3-10x depending on the region. Spike channels

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465750; this version posted October 26, 2021. 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-ND 4.0 International license.

were manually referenced to a local tetrode with no apparent unit firing, bandpass filtered between 200 and 10 kHz, and digitized at 40 kHz. LFP signals were uniformly referenced to ground, amplified 2000x, and bandpass filtered between 0.5Hz and 300 Hz. The animal's position and behavior were tracked via two LEDs fixed to his hyperdrive on one overhead camera, and on one camera positioned behind the two terra-cotta pots each recording at 30 fps. Tetrodes were moved a minimum of 60uM after each recording session to prevent re-sampling the same units. Tetrodes were lowered a minimum of 12 hours before all recording sessions to ensure tetrode stability. IR beam crossings and door movements were automatically logged, and behavioral events were scored manually using video feeds from the two cameras (CinePlex software, Plexon).

Spike sorting and data processing. Spikes were assigned to individual units by offline manual clustering of waveform characteristics (valley amplitude, peak-valley, energy, principal components, and waveform width at half max). Tetrode localization was performed using histology, and further guided by the LFP theta amplitude and phase, and sharp-wave ripple amplitude. Only well-isolated units (Isolation Rating >40 L-Ratio <.001) were kept for further analysis. Behavior was monitored from an overhead camera and a second camera with a close-up view of the test pots. The rats' position was tracked via two light-emitting diodes attached to the implant (one red and one green) and customized MATLAB code. Door movements, and treadmill onset and offset were automatically strobed into the recording amplifier. Test object sampling events were hand-coded using CinePlex editor (Plexon) using video taken at 30 Hz from the camera trained on the reward objects (See Supplemental Video File).

Selecting time cells as input to the hierarchical Bayesian model

As a first pass to identify time cells to be passed on to the hierarchical Bayesian model, we used established methods reported in other papers (Tiganj et al., 2018; Cruzado et al., 2020). This method is conservative in that it uses multiple tests to ensure that a particular unit reliably codes for time across trials and is is meaningfully related to time within the delay period. It does not control for variability across trials nor to does it try to assess the properties of the population.

In order to be classified as a time cell, a unit had to satisfy several criteria. First, to exclude interneurons, we required the firing rate over all delay intervals had to be less than 5 Hz. Second, to be classified as a time cell, units had to show a reliable time field. This was quantified by comparing two models to describe the firing rate as a function of time. One model included a Gaussian-shaped time field, with a mean and a standard deviation, and a constant term. The other model only estimated a constant firing rate. Best-fitting parameters for each model were identified for each unit and each model using scipy's basin-hopping method `scipy.optimize.basinhopping` with truncated Newton algorithm (arguments `stepsize=1000`, `niter=500`, `method=TNC`, `use_jac=T`, `t=1`). In order to pass this criterion, the log-likelihood ratio for the time field model compared to the constant model had to exceed 5.66. The two models differ by three degrees of freedom; 5.66 is the threshold needed to exceed the log-likelihood ratio test with three degrees of freedom at the $p < .01$ level.

To ensure that the units classified as time cells were reliable across trials, we required that the time field model was statistically reliable for both even and odd trials considered

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465750; this version posted October 26, 2021. 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-ND 4.0 International license.

separately. To ensure that the method identified time fields within the delay interval, rather than monotonically ramping or decaying firing (or a time field centered outside the interval), we additionally required that the mean of the Gaussian firing field was within the delay. This method identified 159 units that were classified as time cells which were considered further.

Additionally, the model aims to estimate time fields at the trial level, which will then be pooled to estimate the overall temporal receptive field. To ensure that the variability across trials could be reliably estimated, we required that units considered in the hierarchical Bayesian analysis fired at least one spike on at least 20 trials. This left a total of 131 putative time cells for the hierarchical Bayesian analysis.

Setting the bounds for the power-law distribution. To understand the temporal representation at the population level, we directly estimated the compression level with a power-law distribution. With empirical data, it is often the case that the power-law distribution assumption does not hold for extreme values (too small or too large). With time cells specifically, it is reasonable to assume that the ones fire extremely early or late may be responding to events before or after the delay. Therefore, it is important to find the bounds for power-law distribution if we want an accurate estimation of the compression parameter α. We used the estimated time field peaks from the maximum likelihood method to determine the bounds for power-law distribution for the hierarchical Bayesian model fitting. Clauset et al. (2009) argued that the chosen boundary should make the probability density function from the bounded data as close to the best-fit power-law distribution as possible. Following their advice, we performed Kolmogorov-Smirnov statistic (KS test) to measure the probability that the bounded data was sampled from the best-fit power-law distribution and chose the boundary parameters that produced the highest p-value. For the lower bound parameter (*min* parameter in the hierarchical Bayesian model), we searched the range from 100–400 ms with a 50 ms grid. For the upper bound parameter(*max* parameter in the hierarchical Bayesian model), we searched in the range of 6400–8000 ms with a 400 ms grid (5 percent of the overall trial-length 8000 ms). For each of the lower bound and upper bound pairs, we took the estimated time field peaks that fell in the range and estimated the corresponding power-law distribution (Eq. 4) with the Bayesian model fitting method. The posterior of the compression parameter α is generated through the rStan package with 3 independent Markov-Chain-Monte-Carlo (MCMC) chains (4000 warm-up iterations and 1000 post warm-up samples in each of the MCMC chains). Then we applied the KS test that compares the data against the best-fitting power-law distribution composed with the corresponding α. The cumulative density function (CDF) of the bounded power-law is defined as:

$$P(x) = \frac{x^{1-\alpha} - \min^{1-\alpha}}{\max^{1-\alpha} - \min^{1-\alpha}} \tag{1}$$

In our KS test results, we found the pair with the highest p-value is (350 ms, 7200 ms). We excluded any cells outside of the range (14/159) and applied this range limit for the compression estimation in the Hierarchical Bayesian model fitting below. Although we ended up with one particular boundary pair, the estimated α distribution does not change much between different pairs in our results: after excluding some very early choices (100 ms-250 ms) for lower bound and very late choice (8000 ms), the estimated α posteriors from the remaining pairs are between 0.5 and 1.5 and are centered around 1.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465750; this version posted October 26, 2021. 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-ND 4.0 International license.

Model Time Cells with Hierarchical Bayesian Method

A schematic illustration of the model can be found in figure5. The model generates time field estimation at the individual-trial level. For each trial, the model assumes that the observed spikes are either due to the temporally modulated Gaussian-shaped time field, or some other mechanism that is not temporally variable, which is assumed to be uniformly distributed across the entire trial. Therefore, the probability that cell n fires at any given time point t on trial i is:

$$p_i(t) = a_1\mathcal{N}(\mu_i,\, \sigma_w) + (1 - a_1)\mathcal{U}. \tag{2}$$

where a_1 is the proportion of spikes contributed by the time field and μ_i σ_w are the center and width of the time field, accordingly. Note that although μ_i is specific to trial i, $sigma_w$ is fixed across all trials for the unit.

To quantify the between-trial variability, the model assumes that the time field can shift across trials according to a normal distribution that centers at a parameter M with the standard deviation σ_t:

$$\mu_i \sim \mathcal{N}(M,\, \sigma_t) \tag{3}$$

Finally, we assume that the population of time cells is distributed according to power-law function: $y = x^{-\alpha}$. A probability density function for the time field peak at τ can be derived within a specific range [min, max]:

$$p(M = \tau) = \tau^{-\alpha}/C \tag{4}$$

where

$$C = \int_{\min}^{\max} M^{-\alpha}\, dM \tag{5}$$

is the power-law function integrated over [min, max]. Dividing the power-law function by C ensures the area between [min, max] sums to 1, which makes it a probability distribution within the bounds.

When $\alpha \neq 1$, the result of the integration is:

$$C = \frac{\max^{1-\alpha} - \min^{1-\alpha}}{1 - \alpha} \tag{6}$$

When $\alpha = 1$, the result is:

$$C = \ln(\max) - \ln(\min) \tag{7}$$

As described above, we estimated the range through the Kolmogorov-Smirnov test prior to the hierarchical Bayesian model fitting following the procedure suggested by Clauset et al. (2009).

The value of α reflects the degree of compression in the distribution where a higher α value indicates that a bigger portion of time cells fire early in the delay. Crucially, when α approaches zero, Equation 4 becomes $p(M = \tau) = \frac{1}{\max - \min}$, a uniform distribution between [min,max]. When sorting a group of evenly distributed time cells based on their temporal locations, one would expect the time fields to form a straight line. Another value of great theoretical interest for α is 1: where the time cells are distributed according to logarithmic

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465750; this version posted October 26, 2021. 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-ND 4.0 International license.

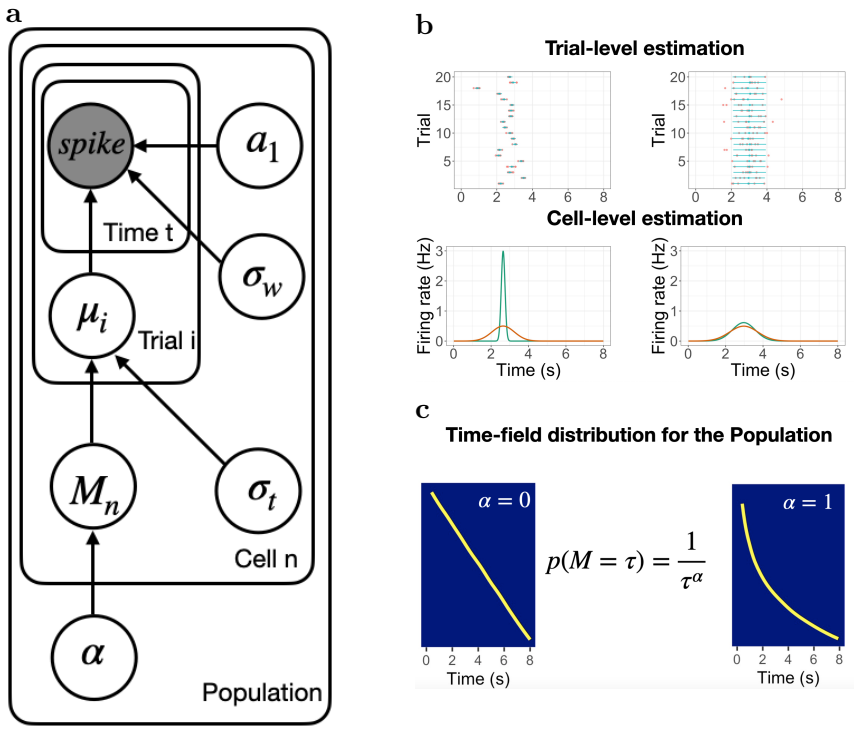

Figure 5. **Schematic illustration of the hierarchical model a.** graphic diagram of the model. Each note represents a variable in the model; filled note represents observed spike-train data and open notes represent latent variables. Arrows represent relationships between variables and plates indicate whether the variable is estimated at the trial level, cell level, or population level. **b-c.** Schematic illustration of model fitting to a group of simulated time cells: the model account for firing pattern at individual trials (top rows in **b**), individual cells (bottom rows in **b**) and the population level **c**. In figure **b** we plot two simulated cells with high (left cell) vs. low (right cell) trial variability. At the trial level, the pink dots indicate the spike train and the blue lines indicate the estimated time fields for each trial. At the cell level, the red lines indicate the average firing patterns across trials and the green lines indicate the estimated time fields for the cell from the model. For both the left and right cells, the overall firing pattern is similar and would results in similar time field width estimation with traditional methods despite the left cell has a much narrower time field. By contrast, fitting at the individual trial level allows the model to estimate the time field width more accurately. In **c**, we plot two hypothetically time field peaks distributions. We plot the distributions of time field peaks (yellow line marks the location of the time field peaks, sorted from early to late) under two specific compression rate α values: time fields are distributed evenly with no compression when $\alpha = 0$ (left panel) vs. time fields are distributed according to logarithmic compression when $\alpha = 1$ (right panel).

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465750; this version posted October 26, 2021. 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-ND 4.0 International license.

compression. When sorting the log-compressed time cells on their temporal locations, one would expect to see more time cells early in the delay. As α increases, for example when α equals 2, the temporal representation becomes more compressed. The bottom panel of Figure 5b shows simulated time cells with these three values of α values.

The posterior distributions of estimated parameters were generated through the Stan package (Stan Development Team, 2021) with 8 independent Markov-Chain-Monte-Carlo (MCMC) chains (`4800 warm-up iterations` and `200 post warm-up samples` in each of the MCMC chains).

Supplementary Materials

S1: Properties of a logarithmic receptor spacing

We suppose that there is a set of receptors characterized by a unimodal receptive field along some continuous variable $x > 0$ corresponding to a physical quantity. Each receptive field is centered on one point along x and is associated with a non-zero width parameter that can change with the center. Let us order the receptive fields by their location along x and refer to the center of the receptive field of the nth receptor as x_n. We say that the receptors form a logarithmically-compressed representation of x if:

$$n = \log_b \left(\frac{x_n}{x_1} \right) \tag{S1}$$

where x_1 is the smallest value of x covered by the set of receptors and the base of the logarithm b controls the spacing between receptors. Note that because the receptors are ordered such that $x_{i+1} > x_i$, this implies that $b > 1$. Equation S1 implies that the spacing between the nth receptor and its near neighbor $n + 1$ is a constant ratio:

$$\frac{x_{n+1}}{x_n} = b \tag{S2}$$

Because the spacing between receptors goes up with x we find

$$\Delta_x \equiv x_{n+1} - x_n = (b - 1)x_n, \tag{S3}$$

. The number of receptors we would expect to find in the neighborhood of a particular value of x is just ${\Delta_x}^{-1}$. This means that if the receptors are sampled without bias, then the probability of finding a receptor near a particular value of x goes down like x^{-1}. That is, if a set of receptors provides a logarithmically-compressed representation of x, then the number density of receptors goes down as a function of x like a power law with exponent -1.

Equation S3, coupled with some minimal arguments about the redundancy of receptors suggests that the width of receptive fields, should go up linearly with x. Let us assume that σ_x, the function relating the width of receptive fields to x is monotonic. We can now consider the ratio $\frac{\sigma_x}{\Delta_x}$ as x grows. If σ_x is monotonic, as x grows without bound, this ratio must either grow without bound, remain constant, or go to zero. If the ratio goes to zero (because σ_x grows more slowly than x) then eventually we would find "gaps" in the coverage of x. If the ratio goes to infinity (because σ_x grows more quickly than x), then this means that the receptive fields at high values of x overlap entirely. In this case the two receptors

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465750; this version posted October 26, 2021. 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-ND 4.0 International license.

are completely redundant with one another. The third alternative is that σ_x is a linear function of x, as is Δ_x. In this case the relative overlap in receptive fields is constant as x grows and adjacent receptors have the same relative overlap in receptive fields throughout x.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465750; this version posted October 26, 2021. 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-ND 4.0 International license.

S2: Quantify How Time Cells Vary across Trials with Model Selection

Because most previous studies ignored the across-trial variability, there is no clear consensus on how the time field may vary across trials. The hierarchical Bayesian model we presented in the main text assumes the width of the time field stays consistent across trials, and the location of the time field shifts from trial to trial (the "location-shift model"). Here we test it against an alternative model where we applied the same hierarchical Bayesian structure with one crucial difference: the time field location is fixed while the width varies across trials (the "width-vary model"). In Fig. S1 we presented example fits from the two models to the same cell. While both models provide similar fits to the overall pattern (lower panel in both figures), the width-vary model is clearly struggled to capture the firing pattern at the individual trial level. When applied the Watanabe-Akaike information criteria (WAIC) test to both fits, we found the majority of units in the population (90/131) prefer the location-shift model. We concluded that the location-shift model is the more accurate way to describe the trial-variability for the population.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465750; this version posted October 26, 2021. 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-ND 4.0 International license.

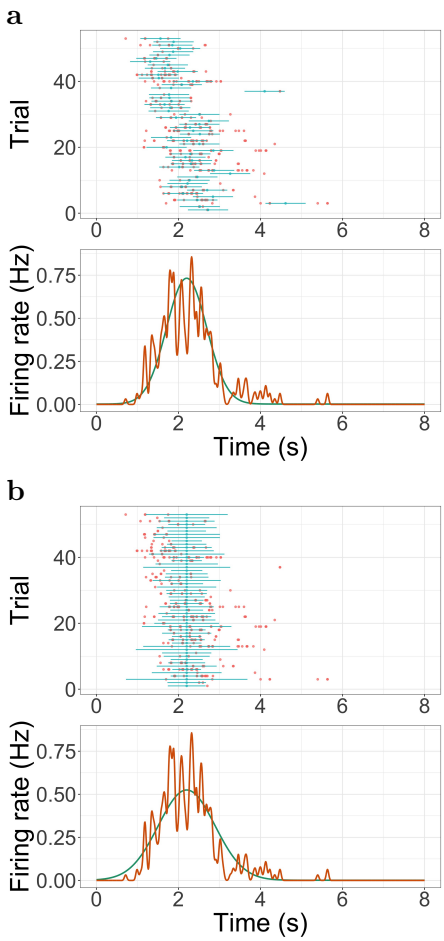

Figure S1. **Example fits from location-shift model (a) and width-vary model (b)** On the top panel of both figures, we plotted fittings to individual trials (in blue lines, width indicate time field width for the trial) and raster (pink dots). On the bottom panel of both figures, we plotted the estimated time field (green line) and the average firing pattern (red line) of the unit.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465750; this version posted October 26, 2021. 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-ND 4.0 International license.

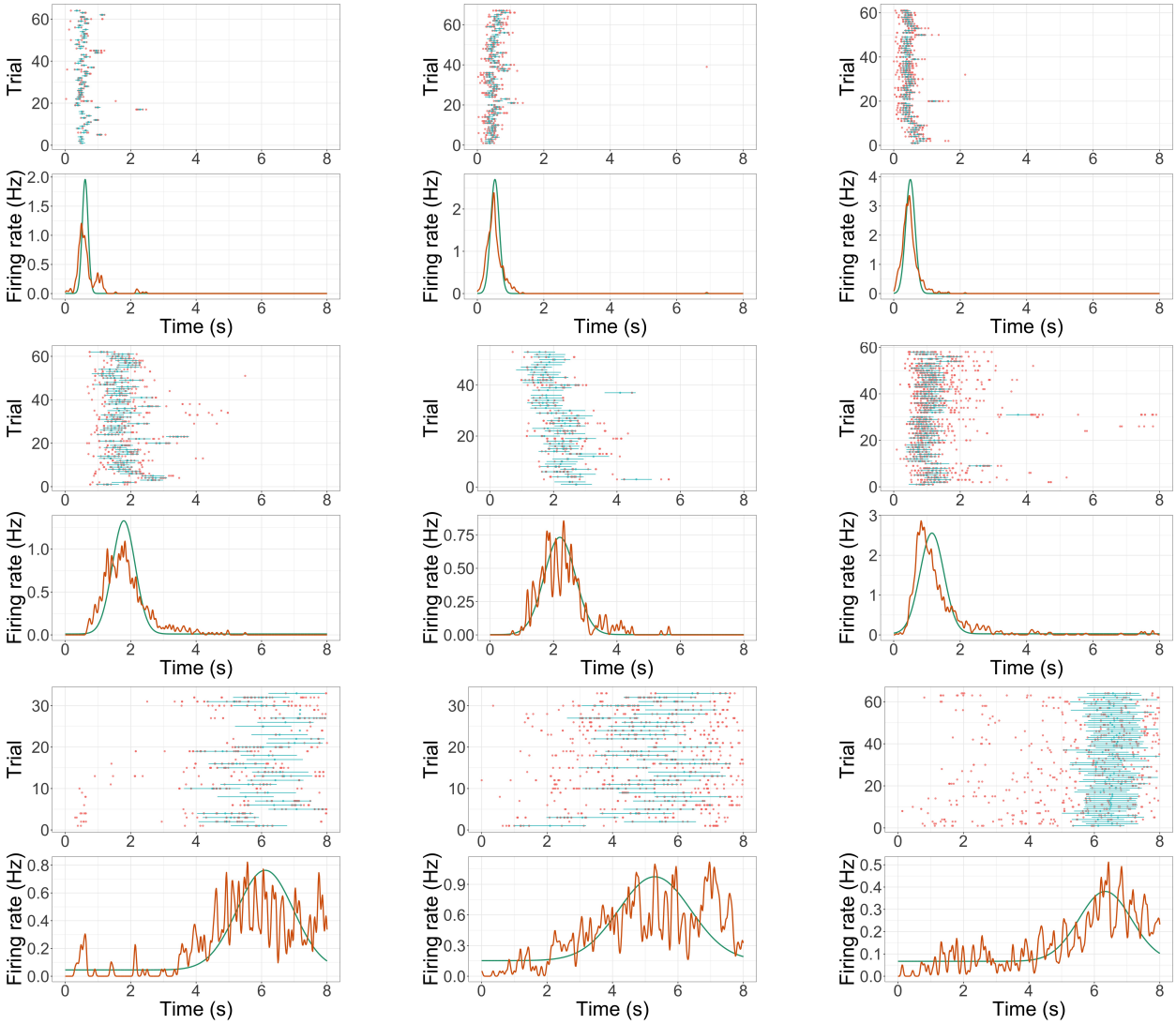

Figure S2. **Additional example time cell fits.** Format is the same as Figure 2a.

S3: Additional Model Fits

We provided additional example time cells in figure S2

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465750; this version posted October 26, 2021. 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-ND 4.0 International license.

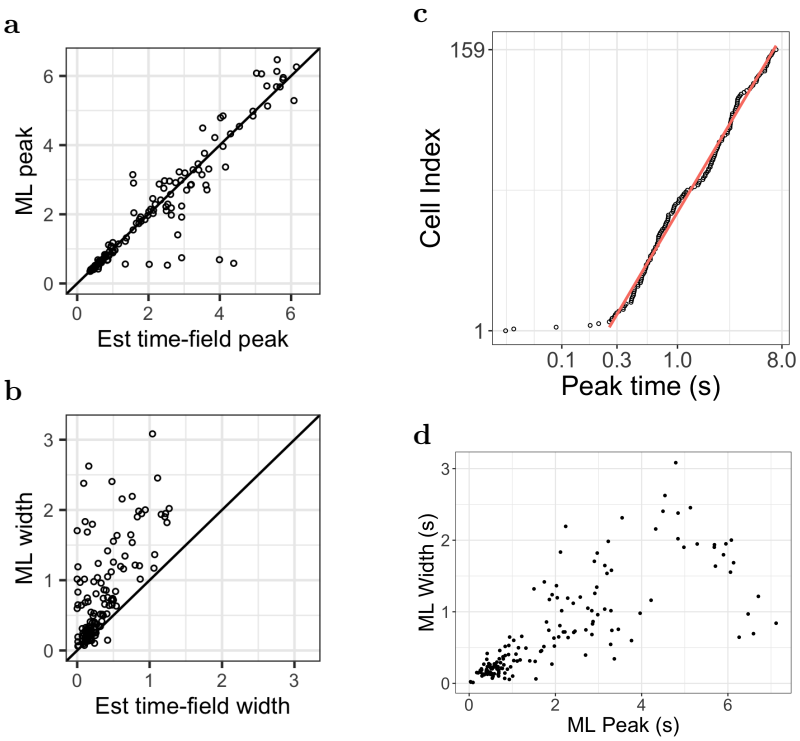

Figure S3. **Additional analysis comparing the maximum likelihood fits and the hierarchical Bayesian fits a-b** Maximum likelihood estimated time field peaks (a) and width (b) as a function of their corresponding estimation from the Hierarchical Bayesian method (x-axis). Points on the diagonal line are units where the two values are the same. Points above the diagonal line are units the higher values in the maximum likelihood fits. **c** Maximum likelihood estimated time field peaks plotted on a log-scale (x-axis), sorted on the order of their peak time location (y-axis). Pink line represents linear fit on log data excluding extremely early time cells (prior to 0.25 s). **d** Maximum likelihood estimated time field widths as a function of their time field peaks.

S4: Analyze Maximum Likelihood Fits

In this section, we reported the results from Maximum likelihood fits and compared them with results from Hierarchical Bayesian. In Fig. S3a, we compare the Maximum likelihood estimated time field peaks as a function of hierarchical Bayesian estimation. As shown in the figure, aside from a few outliers cluster toward the bottom, most of the cells are closely alongside the diagonal line, which suggests that the results from both methods are close to each other. Similarly, when plotted on the log scale, the sorted time field peaks from Maximum likelihood methods tiles evenly across the log scale (Fig. S3c). This further demonstrated the robustness of the compression rate estimation. In Fig. S3b, we plotted the maximum likelihood estimated time field width as a function of hierarchical Bayesian estimation. Contrary to the plot of estimation time field peaks, most of the cells here are above the diagonal line, which means the maximum likelihood estimated time field widths tend to be larger than the ones from the hierarchical Bayesian model. Paired t-test results confirmed that the difference is significant ($t(131) = 9.38$, $p < 0.001$, 95 percent confidence interval: [0.34,0.52]). The result is expected as the maximum likelihood estimation does

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465750; this version posted October 26, 2021. 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-ND 4.0 International license.

not separate the time field width from the across-trial variability. As reported in the main text, there is a significant linear relationship between hierarchical Bayesian estimated time field widths and time field peaks. However, When plot the Maximum likelihood estimated time field widths as a function of their time field peaks, linear regression analysis shows that both linear (0.64 ± 0.06, $p < 0.001$) and quadratic term (-0.06 ± 0.01, $p < 0.001$) are significant ($R^2 = 0.674$, Fig. S3d). Again it is likely due to the inclusion of trial-variability in the time field width estimation.

Acknowledgments. The authors gratefully acknowledge support from ONR MURI N00014-16-1-2832, NIBIB R01EB022864, NIMH R01MH112169, and NIMH R01MH095297. The authors gratefully acknowledge the contributions of Howard Eichenbaum to designing the behavioral task and data collection in this study.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465750; this version posted October 26, 2021. 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-ND 4.0 International license.

References

Arzy, S., Adi-Japha, E., & Blanke, O. (2009). The mental time line: An analogue of the mental number line in the mapping of life events. *Consciousness and cognition*, *18*(3), 781–785.

Bellmund, J. L., Deuker, L., & Doeller, C. F. (2019). Mapping sequence structure in the human lateral entorhinal cortex. *eLife*, *8*, e45333.

Bjerknes, T. L., Dagslott, N. C., Moser, E. I., & Moser, M.-B. (2018). Path integration in place cells of developing rats. *Proceedings of the National Academy of Sciences*, *115*(7), E1637–E1646.

Brannon, E. M. (2006). The representation of numerical magnitude. *Current opinion in neurobiology*, *16*(2), 222–229.

Bright, I. M., Meister, M. L. R., Cruzado, N. A., Tiganj, Z., Buffalo, E. A., & Howard, M. W. (2020). A temporal record of the past with a spectrum of time constants in the monkey entorhinal cortex. *Proceedings of the National Academy of Sciences*, *117*, 20274-20283.

Buzsáki, G., & Mizuseki, K. (2014). The log-dynamic brain: how skewed distributions affect network operations. *Nature Reviews Neuroscience*, *15*(4), 264–278.

Cai, D. J., Aharoni, D., Shuman, T., Shobe, J., Biane, J., Song, W., ... Silva, A. (2016). A shared neural ensemble links distinct contextual memories encoded close in time. *Nature*, *534*(7605), 115–118.

Cantlon, J. F., Brannon, E. M., Carter, E. J., & Pelphrey, K. A. (2006). Functional imaging of numerical processing in adults and 4-y-old children. *PLoS biology*, *4*(5), e125.

Clauset, A., Shalizi, C. R., & Newman, M. E. (2009). Power-law distributions in empirical data. *SIAM review*, *51*(4), 661–703.

Cohen, J. E., Ross, R. S., & Stern, C. E. (2018). Predictability matters: role of the hippocampus and prefrontal cortex in disambiguation of overlapping sequences. *Learning & Memory*, *25*(8), 335–346.

Cruzado, N. A., Tiganj, Z., Brincat, S. L., Miller, E. K., & Howard, M. W. (2020). Conjunctive representation of what and when in monkey hippocampus and lateral prefrontal cortex during an associative memory task. *Hippocampus*, *30*, 1332-1346.

Daniel, P. M., & Whitteridge, D. (1961). The representation of the visual field on the cerebral cortex in monkeys. *Journal of Physiology*, *159*, 203-21.

Dannenberg, H., Kelley, C., Hoyland, A., Monaghan, C. K., & Hasselmo, M. E. (2019). The firing rate speed code of entorhinal speed cells differs across behaviorally relevant time scales and does not depend on medial septum inputs. *Journal of Neuroscience*, 1450–18.

Dehaene, S., & Brannon, E. (2011). *Space, time and number in the brain: Searching for the foundations of mathematical thought*. Academic Press.

Deuker, L., Bellmund, J. L., Schröder, T. N., & Doeller, C. F. (2016). An event map of memory space in the hippocampus. *eLife*, *5*, e16534.

Dumoulin, S. O., & Wandell, B. A. (2008). Population receptive field estimates in human visual cortex. *Neuroimage*, *39*(2), 647–660.

Eagleman, D. M. (2008). Human time perception and its illusions. *Current opinion in neurobiology*, *18*(2), 131–136.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465750; this version posted October 26, 2021. 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-ND 4.0 International license.

Egorov, A. V., Hamam, B. N., Fransén, E., Hasselmo, M. E., & Alonso, A. A. (2002). Graded persistent activity in entorhinal cortex neurons. *Nature*, *420*(6912), 173-8.

Eichenbaum, H. (2017). On the integration of space, time, and memory. *Neuron*, *95*(5), 1007-1018. doi: 10.1016/j.neuron.2017.06.036

Fechner, G. T. (1860). *Elemente der psychophysik* (Vol. 2). Breitkopf u. Härtel.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, *8*(7), 307–314.

Folkerts, S., Rutishauser, U., & Howard, M. (2018). Human episodic memory retrieval is accompanied by a neural contiguity effect. *Journal of Neuroscience*, *38*, 4200-4211.

Gallistel, C. R. (1989). Animal cognition: the representation of space, time and number. *Annual Review of Psychology*, *40*, 155-89. doi: 10.1146/annurev.ps.40.020189.001103

Gallistel, C. R., & Gelman, R. (2000). Non-verbal numerical cognition: From reals to integers. *Trends in Cognitive Sciences*, *4*(2), 59–65.

Gattass, R., Gross, C., & Sandell, J. (1981). Visual topography of v2 in the macaque. *Journal of Comparative Neurology*, *201*(4), 519–539.

Gibbon, J., Church, R. M., & Meck, W. H. (1984). Scalar timing in memory. *Annals of the New York Academy of sciences*, *423*(1), 52–77.

Glezer, V. (1965). The receptive fields of the retina. *Vision research*, *5*(10-11), 497–525.

Gothard, K. M., Hoffman, K. L., Battaglia, F. P., & McNaughton, B. L. (2001). Dentate gyrus and CA1 ensemble activity during spatial reference frame shifts in the presence and absence of visual input. *Journal of Neuroscience*, *21*(18), 7284-92.

Gothard, K. M., Skaggs, W. E., Moore, K. M., & McNaughton, B. L. (1996). Binding of hippocampal CA1 neural activity to multiple reference frames in a landmark-based navigation task. *Journal of Neuroscience*, *16*(2), 823-35.

Guo, C., Huson, V., Macosko, E. Z., & Regehr, W. G. (2021). Graded heterogeneity of metabotropic signaling underlies a continuum of cell-intrinsic temporal responses in unipolar brush cells. *Nature Communications*, *12*(1), 1–12.

Harland, B., Contreras, M., Souder, M., & Fellous, J.-M. (2021). Dorsal ca1 hippocampal place cells form a multi-scale representation of megaspace. *Current Biology*.

Healey, M. K., & Kahana, M. J. (2013). Memory search is governed by universal principlesÂŮnot idiosyncratic strategies. *Journal of Experimental Psychology: General.*.

Howard, M. W., & Eichenbaum, H. (2013). The hippocampus, time, and memory across scales. *Journal of Experimental Psychology: General*, *142*(4), 1211-30. doi: 10.1037/a0033621

Howard, M. W., & Kahana, M. J. (1999). Contextual variability and serial position effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 923-941.

Howard, M. W., Shankar, K. H., Aue, W., & Criss, A. H. (2015). A distributed representation of internal time. *Psychological Review*, *122*(1), 24-53.

Howard, M. W., Youker, T. E., & Venkatadass, V. (2008). The persistence of memory: Contiguity effects across several minutes. *Psychonomic Bulletin & Review*, *15*(PMC2493616), 58-63.

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), 1165–1178.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465750; this version posted October 26, 2021. 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-ND 4.0 International license.

Hubel, D. H., & Wiesel, T. N. (1974). Uniformity of monkey striate cortex: a parallel relationship between field size, scatter, and magnification factor. *Journal of Comparative Neurology*, *158*(3), 295-305. doi: 10.1002/cne.901580305

Jacques, B. G., Tiganj, Z., Sarkar, A., Howard, M. W., & Sederberg, P. B. (2021). *Sithcon: A neural network robust to variations in input scaling on the time dimension.*

Jin, D. Z., Fujii, N., & Graybiel, A. M. (2009). Neural representation of time in cortico-basal ganglia circuits. *Proceedings of the National Academy of Sciences*, *106*(45), 19156–19161.

Kahana, M. J. (2012). *Foundations of human memory.* OUP USA.

Kahana, M. J., Howard, M., & Polyn, S. (2008). Associative processes in episodic memory. In H. L. Roediger III (Ed.), *Cognitive psychology of memory, Vol. 2 of learning and memory - a comprehensive reference (J. Byrne, Editor)* (p. 476-490). Oxford: Elsevier.

Keene, C. S., Bladon, J., McKenzie, S., Liu, C. D., O'Keefe, J., & Eichenbaum, H. (2016). Complementary functional organization of neuronal activity patterns in the perirhinal, lateral entorhinal, and medial entorhinal cortices. *Journal of Neuroscience*, *36*(13), 3660–3675.

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

Lee, M. D. (2011). How cognitive modeling can benefit from hierarchical bayesian models. *Journal of Mathematical Psychology*, *55*(1), 1–7.

Liu, Y., Levy, S. J., Mau, W., Geva, N., Rubin, A., Ziv, Y., ... Howard, M. (2021). Consistent population activity on the scale of minutes in the mouse hippocampus. *bioRxiv*.

Liu, Y., Tiganj, Z., Hasselmo, M. E., & Howard, M. W. (2019). A neural microcircuit model for a scalable scale-invariant representation of time. *Hippocampus*, *29*(3), 260–274.

Logan, G. D. (2021). Serial order in perception, memory, and action. *Psychological Review*, *128*(1), 1.

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., Lepage, K. Q., Eden, U. T., & Eichenbaum, H. (2011). Hippocampal "time cells" bridge the gap in memory for discontiguous events. *Neuron*, *71*(4), 737-749.

Mack, C. C., Cinel, C., Davies, N., Harding, M., & Ward, G. (2017). Serial position, output order, and list length effects for words presented on smartphones over very long intervals. *Journal of Memory and Language*, *97*, 61–80.

Manning, J. R., Polyn, S. M., Litt, B., Baltuch, G., & Kahana, M. J. (2011). Oscillatory patterns in temporal lobe reveal context reinstatement during memory search. *Proceedings of the National Academy of Science, USA*, *108*(31), 12893-7.

Mau, W., Hasselmo, M. E., & Cai, D. J. (2020). The brain in motion: How ensemble fluidity drives memory-updating and flexibility. *Elife*, *9*, e63550.

Mau, W., Sullivan, D. W., Kinsky, N. R., Hasselmo, M. E., Howard, M. W., & Eichenbaum, H. (2018). The same hippocampal CA1 population simultaneously codes temporal

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465750; this version posted October 26, 2021. 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-ND 4.0 International license.

information over multiple timescales. *Current Biology*, *28*, 1499-1508.

Mello, G. B., Soares, S., & Paton, J. J. (2015). A scalable population code for time in the striatum. *Current Biology*, *25*(9), 1113–1122.

Moore, B. C. (2012). *An introduction to the psychology of hearing*. Brill.

Nieder, A., & Dehaene, S. (2009). Representation of number in the brain. *Annual Review of Neuroscience*, *32*, 185-208. doi: 10.1146/annurev.neuro.051508.135550

Nieder, A., & Miller, E. K. (2003). Coding of cognitive magnitude: compressed scaling of numerical information in the primate prefrontal cortex. *Neuron*, *37*(1), 149-57.

Nielson, D. M., Smith, T. A., Sreekumar, V., Dennis, S., & Sederberg, P. B. (2015). Human hippocampus represents space and time during retrieval of real-world memories. *Proceedings of the National Academy of Sciences*, *112*(35), 11078–11083.

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

Peer, M., Salomon, R., Goldberg, I., Blanke, O., & Arzy, S. (2015). Brain system for mental orientation in space, time, and person. *Proceedings of the National Academy of Sciences*, *112*(35), 11072–11077.

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review*, *116*, 129-156.

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.

Robinson, N. T., Priestley, J. B., Rueckemann, J. W., Garcia, A. D., Smeglin, V. A., Marino, F. A., & Eichenbaum, H. (2017). Medial entorhinal cortex selectively supports temporal coding by hippocampal neurons. *Neuron*, *94*(3), 677–688.

Rubin, A., Geva, N., Sheintuch, L., & Ziv, Y. (2015). Hippocampal ensemble dynamics timestamp events in long-term memory. *eLife*, *4*, e12247.

Schapiro, A. C., Turk-Browne, N. B., Norman, K. A., & Botvinick, M. M. (2016). Statistical learning of temporal community structure in the hippocampus. *Hippocampus*, *26*(1), 3–8.

Schwartz, E. L. (1980). Computational anatomy and functional architecture of striate cortex: a spatial mapping approach to perceptual coding. *Vision Research*, *20*(8), 645-69.

Sederberg, P. B., Howard, M. W., & Kahana, M. J. (2008). A context-based theory of recency and contiguity in free recall. *Psychological Review*, *115*, 893-912.

Sheehan, D. J., Charczynski, S., Fordyce, B. A., Hasselmo, M. E., & Howard, M. W. (2021). A compressed representation of spatial distance in the rodent hippocampus. *bioRxiv*.

Shikano, Y., Ikegaya, Y., & Sasaki, T. (2021). Minute-encoding neurons in hippocampal-striatal circuits. *Current Biology*, *31*(7), 1438–1449.

Soares, S., Atallah, B. V., & Paton, J. J. (2016). Midbrain dopamine neurons control judgment of time. *Science*, *354*(6317), 1273-1277. doi: 10.1126/science.aah5234

Stensola, H., Stensola, T., Solstad, T., Frøland, K., Moser, M.-B., & Moser, E. I. (2012). The entorhinal grid map is discretized. *Nature*, *492*(7427), 72–78.

Talmi, D., Lohnas, L. J., & Daw, N. D. (2019). A retrieved context model of the emotional modulation of memory. *Psychological Review*, *126*(4), 455.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.25.465750; this version posted October 26, 2021. 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-ND 4.0 International license.

Taxidis, J., Pnevmatikakis, E. A., Dorian, C. C., Mylavarapu, A. L., Arora, J. S., Samadian, K. D., ... Golshani, P. (2020). Differential emergence and stability of sensory and temporal representations in context-specific hippocampal sequences. *Neuron*.

Tiganj, Z., Cromer, J. A., Roy, J. E., Miller, E. K., & Howard, M. W. (2018). Compressed timeline of recent experience in monkey lPFC. *Journal of Cognitive Neuroscience*, *30*, 935-950.

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., Shankar, K. H., & Howard, M. W. (2017). Scale invariant value computation for reinforcement learning in continuous time. In *AAAI 2017 Spring Symposium Series - Science of Intelligence: Computational Principles of Natural and Artificial Intelligence.*

Tsao, A., Sugar, J., Lu, L., Wang, C., Knierim, J. J., Moser, M.-B., & Moser, E. I. (2018). Integrating time from experience in the lateral entorhinal cortex. *Nature*, *561*, 57-62.

Tubridy, S., & Davachi, L. (2011). Medial temporal lobe contributions to episodic sequence encoding. *Cerebral Cortex*, *21*(2), 272-80. doi: 10.1093/cercor/bhq092

Tulving, E., & Madigan, S. A. (1970). Memory and verbal learning. *Annual Review of Psychology*, *21*, 437-484.

Uitvlugt, M. G., & Healey, M. K. (2019). Temporal proximity links unrelated news events in memory. *Psychological science*, *30*(1), 92–104.

Unsworth, N. (2008). Exploring the retrieval dynamics of delayed and final free recall: Further evidence for temporal-contextual search. *Journal of Memory and Language*, *59*, 223-236.

Van Essen, D. C., Newsome, W. T., & Maunsell, J. H. (1984). The visual field representation in striate cortex of the macaque monkey: asymmetries, anisotropies, and individual variability. *Vision Research*, *24*(5), 429-48.

Wei, X.-X., Prentice, J., & Balasubramanian, V. (2015). A principle of economy predicts the functional architecture of grid cells. *Elife*, *4*, e08362. doi: 10.7554/eLife.08362

Yaffe, R. B., Kerr, M. S. D., Damera, S., Sarma, S. V., Inati, S. K., & Zaghloul, K. A. (2014). Reinstatement of distributed cortical oscillations occurs with precise spatiotemporal dynamics during successful memory retrieval. *Proceedings of the National Academy of Sciences*, *111*(52), 18727-32. doi: 10.1073/pnas.1417017112

Ziv, Y., Burns, L. D., Cocker, E. D., Hamel, E. O., Ghosh, K. K., Kitch, L. J., ... Schnitzer, M. J. (2013). Long-term dynamics of CA1 hippocampal place codes. *Nature Neuroscience*, *16*(3), 264-6. doi: 10.1038/nn.3329