# Scanning along a compressed timeline of the future

Inder Singh and Marc W. Howard
Center for Memory and Brain
Department of Psychological and Brain Sciences
Boston University

**Abstract**

Several authors have suggested a deep symmetry between the psychological processes that underlie our ability to remember the past and make predictions about the future. The judgment of recency (JOR) task measures temporal order judgments for the past by presenting pairs of probe stimuli; participants choose the probe that was presented more recently. We performed a short-term relative JOR task and introduced a novel judgment of imminence (JOI) task to study temporal order judgments for the future. In the JOI task, participants were trained on a probabilistic sequence. During a test phase, the sequence was occasionally interrupted with pairs of probes. Participants chose the probe that they expected would be presented sooner. Replicating prior work, we found that in JOR the correct RT depended only on the recency of the more recent probe. This suggests that memory for the past was supported by a backward self-terminating search model operating on a temporally-organized representation of the past. Analogously, in the JOI task we find that correct RT depended only on the imminence of the more imminent probe. By analogy to the JOR results, this suggests a forward self-terminating search model operating on a temporally-organized representation of the future. Critically, in both JOR and JOI the increase in RT with recency/imminence was sublinear, suggesting that both the timeline for the future and the timeline of the past are compressed. These results place strong constraints on computational models for constructing the predicted future.

## Introduction

Many authors have hypothesized a deep symmetry between the ability to remember the past and imagine the future. In episodic memory in particular, the ability to remember

---

Address correspondence to `inders@bu.edu` or `marc777@bu.edu`. Inder Singh's current address is the 125 NI, Department of Psychology, Northeastern University. Supported by NSF IIS 1631460 and NIBIB R01EB022864.

episodes from the past has been linked to the ability to simulate the future (Tulving, 1985). Recent findings from neuroimaging studies have found some support for this view (e.g., Schacter, Addis, & Buckner, 2007; Bar, 2009). Deficits in the ability to predict the future are known to be comorbid with deficits in episodic memory due to aging (Addis, Wong, & Schacter, 2007) and neurological insults (Hassabis, Kumaran, Vann, & Maguire, 2007). Some authors have even argued that optimal prediction could indeed be the guiding principle underlying the functional organization of the brain (Friston, 2010; Bialek, 2012; Palmer, Marre, Berry, & Bialek, 2015). While many authors have suggested a symmetry between the mechanisms that allow us to access the past and the future, there is a dearth of behavioral paradigms that allow one to directly compare memory access for past and future events under controlled laboratory conditions.

In this paper, we introduce a future-time analog of the relative judgment of recency (JOR) task. By analogy to the way the JOR evaluates participants' ability to judge the relative time at which past events occurred, this new paradigm tests the ability of participants' ability to judge the imminence of future events over the scale of a few seconds. The classic finding from short-term relative JOR tasks is that correct response time (RT) depends on the recency of the more recent probe but not the recency of the less recent probe (Muter, 1979; Hacker, 1980; Hockley, 1984). For decades, researchers have argued that this finding is consistent with a self-terminating backward scan along a temporally-organized memory representation (Muter, 1979; Hacker, 1980; Hockley, 1984; McElree & Dosher, 1993). Because the scan starts at the present and proceeds backwards in time, RTs show a recency effect. Because the scan is self-terminating this naturally accounts for the finding that correct RTs do not depend on the recency of the less-recent probe. This backward scanning account also naturally explains the finding that incorrect RTs depend on the recency of the (incorrectly) selected probe item under the assumption that errors occur because the scan continues past the correct probe.

A backward scanning model implies that memory for the list is organized along a temporal axis; in order for memory to be sequentially "scanned" it must be temporally organized. This is consistent with longstanding theories of human memory (James, 1890; Brown, Neath, & Chater, 2007; Balsam & Gallistel, 2009; Howard, Shankar, Aue, & Criss, 2015) which argue further that in memory the temporal dimension coding for the past should be compressed (Brown et al., 2007; Howard et al., 2015). A serial scan of a compressed temporal representation should result in a sublinear increase in RTs, rather than a linear increase in RTs as one would expect from scanning of an uncompressed representation of time. [1]

If there is a symmetry between memory for the past and prediction of the future, we would expect analogous results to hold when participants are asked to make judgments about the expected time of anticipated future events. The relative JOR task requires the participant to select the probe item from a past list that was presented closer to the present (Figure 1b). We introduce a novel judgment of imminence (JOI) paradigm that asks participants to select the future probe item that *is anticipated* closer to the present (Figure 1c). The task is closely analogous to the JOR task so that the time to access a memory and an anticipated future event can be directly compared.

[1] Although previous results are consistent with a sublinear increase in RT with recency (see e.g., Fig. 7 of Hockley, 1984, Fig. 5 Hacker, 1980) this was not statistically evaluated in those prior studies.

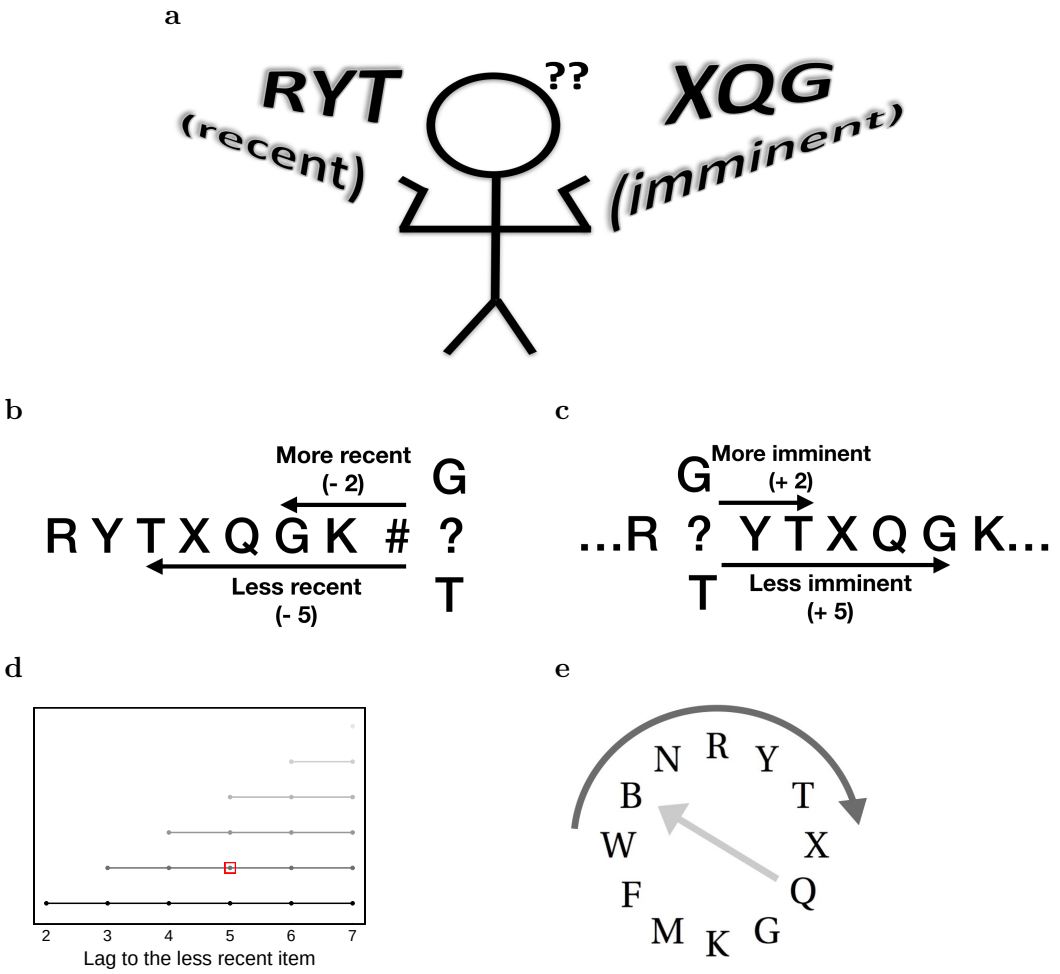


*Figure 1.* **Scanning in the judgment of recency (JOR) and judgment of imminence (JOI) task. a.** A hypothesis for an ordered representation for the past and the future. Representations of both the past and the future are stored along a temporal dimension that can be scanned. The spacing between the letters is foreshortened to suggest compression of the temporal dimension. **b.** In JOR, the participants are shown a list of letters (RYT. . . ) followed by a probe containing two letters from the list (G and T). Participants choose the probe item that was experienced more recently. **c.** In JOI, participants learn a probabilistic sequence. After learning these transition probabilities, the sequence is interrupted by a probe containing two letters from the list (G and T). Participants choose the probe that is more imminent. **d.** The two letters in a probe correspond to a more recent/imminent probe and a less recent/imminent probe. In the RT plots that follow, the shading represents the lag to the more recent item and the less recent item is plotted on the x-axis. The red square represents the RT corresponding to the probe (G and T) in the JOR task. **e.** The JOI sequence consists of 13 consonants that are arranged in a ring. Participants are shown the next item in the sequence with a probability of 0.9 and with a probability of 0.1 any other item in the list is shown and so on.

## Results

To summarize the results, correct RT in both JOR and JOI depended strongly on the recency/imminence of the more recent/imminent probe, consistent with a scanning model that examines a temporal representation starting from the present. Moreover, the effect of recency/imminence on RT was sublinear suggesting that the representation being scanned is compressed.

### *Results JOR*

Accuracy in the JOR was consistent with prior studies of short-term JOR. Briefly, there was a recency effect on accuracy such that more recent probes were more accurate and also a distance effect such that participants were more accurate when there was a larger difference between the recency of the two probes (see Supplementary Information for details).

*Correct RT depended only on the recency of the selected probe.* The median RTs for the correct responses depended strongly on the recency of the more recent (correct) probe as seen in Figure 2a. RT varied from $.72 \pm .02$ s for the most recent probe to $1.36 \pm .06$ s for a recency of $-6$. In contrast to the distance effect seen in accuracy Figure 2a, the lines in Figure 2a appear to be flat. In order to assess this distance effect more directly, we calculated the slopes of lines in Figure 2a separately for each participant and performed a Bayesian t-test (Rouder, Speckman, Sun, Morey, & Iverson, 2009) on the slopes. This analysis showed "substantial evidence" (Wetzels & Wagenmakers, 2012; Kass & Raftery, 1995; Jeffreys, 1998) favoring the hypothesis that the slopes are not different from 0 (JZS Bayes Factor = 3.3). A linear mixed effects analysis allowing for independent intercepts for each participant showed a significant effect of the recency to the more recent probe, $.124 \pm .006$ s, $t(478) = 21.6$, $p < 0.001$. These results replicate prior studies, but extend them by establishing positive evidence for the null distance effect using the Bayesian t-test.

*RT varied sub-linearly with recency of the selected probe.* The separation between the nearly-horizontal lines in Figure 2a indicate that correct RTs depended prominently on the recency to the more recent item. Further it appears that the spacing between these lines goes down as the recency increases. This suggests that the RT depends sub-linearly on the recency to the more recent probe, as predicted by a backward self-terminating scanning model that scans along a temporally-compressed representation.

In order to evaluate whether scanning times depended on recency to the more recent probe in a sublinear fashion, two models were compared. In one model, RT was regressed onto the recency of the more recent probe. In the other RT was regressed onto the logarithm of the more recent recency. The log model fit better than the linear model, $\Delta AIC = 4.1$ (log model is 60.34 times more likely as compared to the linear model). A linear mixed effects analysis allowing for independent intercepts for each participant showed a significant effect of the $log_2$ recency of the more recent probe, $.24 \pm .01$ s, $t(478) = 21.74$, $p < 0.001$.

As an additional test for sublinearity the linear model was compared to polynomial models with various powers of recency. A quadratic model fit better than the linear model $\Delta AIC = 3.7$. The quadratic term had a negative regression coefficient $-0.6 \pm 0.2$, $t(477) = -2.6$, $p < 0.01$. Including higher order terms did not further improve the fit. Consistent

with the conclusions of the logarithmic analysis reported above, both approaches found evidence that the effect the recency to the more recent probe on correct RT was sublinear. The finding of sublinearity is consistent with prior studies (Hacker, 1980; Hockley, 1984). However to our knowledge it had not previously been statistically evaluated.

*Results JOI*

Accuracy in the JOI task depended on the imminence of the more imminent probe; we did not observe a distance effect. Accuracy was higher than in the JOR task. Findings related to accuracy are described in more detail in the Supplementary Information.

*Correct RT depended strongly on the imminence of the more imminent probe, but not on the imminence of the less imminent probe* . Correct RTs depended strongly on the more imminent item, as can be seen by the separation of the nearly-horizontal lines in Figure 2b. Correct RT varied from $1.01 \pm .05$ s for the most imminent item to $1.81 \pm .05$ s for an imminence of five. The lines in Figure 2b are nearly flat, suggesting the absence of a distance effect. In order to assess this more directly, the slopes of the lines in Figure 2b were calculated separately for each participant and a Bayesian t-test (Rouder et al., 2009) was performed on the slopes. This analysis showed "substantial evidence" (Wetzels & Wagenmakers, 2012) favoring the hypothesis (JZS Bayes Factor = 9.07) that the slopes are not different from 0. This suggest that the RT depended prominently on the imminence to the more imminent item.

A linear mixed effects analysis allowing for independent intercepts for each participant showed a significant effect of the imminence of the more imminent probe, $.195 \pm .009$ s, $t(227) = 24.1$, $p < 0.001$.

*RT varied sub-linearly with the imminence of the selected probe* . Upon visual inspection, Figure 2b suggests that the spacing between the nearly-horizontal lines goes down as the imminence decreases. This implies that correct RT seems to get depend sub-linearly on the imminence of the more imminent probe.

In order to test this hypothesis, two models were compared. In one model median RT was regressed on imminence directly. In the other model median RT was regressed onto the logarithm of imminence. The log model fit better than the linear model, $\Delta AIC = 32.6$ (log model is $10^7$ times more likely as compared to the linear model). A linear mixed effects analysis allowing for independent intercepts for each participant showed a significant effect of the $log_2$ imminence of the more imminent probe, $.34 \pm .01$ s, $t(227) = 26.5$, $p < 0.001$.

As an additional test, we also ran a separate analysis comparing polynomial fits to the various powers of the imminence of the more imminent item. The quadratic model provided a better fit than the linear model $\Delta log\ likelihood = 26.6$ with a negative slope of the quadratic term $-1.0 \pm 0.2$, $t(226) = -5.7$, $p < 0.01$. The quadratic model also fit better than models with higher order polynomials. Thus RT appears to be temporally compressed in short term memory. This result is consistent with the conclusions of the logarithmic analysis reported above. Both approaches found evidence that the effect the imminence of the more imminent probe on correct RT was sub-linear, as predicted by a scanning model that scans along a temporally-compressed representation.

## Discussion

The goal of this study was to develop a paradigm for making predictions in prospective memory that is analogous to relative JOR task used to study memory for the past. This study found that prospective memory for the future has the same properties as memory for the past. The pattern of results for both memory for the past and prediction of the future are consistent with scanning models of memory that posit that information is stored in a temporally organized representation that can be sequentially accessed in order to make a response (Murdock, 1974). The major difference between memory for the past and prediction of the future is the direction of the scan. In both cases, the scan extends away from the present; towards the past in memory but forwards towards the future in prediction.

Critically, scanning of both the past and the future suggests that the temporal representation is compressed, with decreasing accuracy for events further from the present. This compression is reflected in the sublinear increase of correct RT with distance from the present (Figure 2c,d). In both cases we can certainly reject the hypothesis that RT increases linearly. The results are consistent with a number of forms of compression and we cannot make a strong conclusion about the precise quantitative form of the compression. However, the results are at least consistent with a logarithmic form of compression. Logarithmic compression is consistent with the Weber-Fechner law and implements a scale-invariant temporal scale, as predicted by quantitative models of memory (Brown et al., 2007; Howard et al., 2015).

The similarity of the results for memory for the past and expectation of the future are more striking when one notes the numerous procedural differences between the task. First, the lists were presented at a much faster rate in the JOR experiment than in the JOI experiment. The participants in the JOI had only a single list to remember throughout the session. In contrast, in the JOR task participants learned many lists using the same stimuli over the course of a session. Moreover, participants in the JOR tasks have a well defined expectation on when the probe will be shown since it is always presented at the end of the list. In the JOI task however, the probe was presented probabilistically. In light of these procedural differences between the JOR task and the JOI task, the similarity of the core results is even more striking. The overall slope of the median RT with the $log_2$ recency/imminence in JOI and JOR are $.34 \pm .01$ and $24 \pm .01$. Despite these procedural differences, the scanning rates are comparable.

### *These data represent a challenge for RL-based models of the future*

How could the brain construct a compressed timeline for future events? It is clear that in order to predict future events, participants must build up some understanding of the statistical relationships between the stimuli. Reinforcement learning models have been extremely influential in the cognitive neuroscience of prediction (Schultz, Dayan, & Montague, 1997; Daw & Dayan, 2014). In this view, the brain can predict the future using either a model-free or a model-based system; these two systems have different properties and can be brought to bear under appropriate circumstances (e.g., Daw, Niv, & Dayan, 2005; Russek, Momennejad, Botvinick, Gershman, & Daw, 2017; Momennejad et al., 2017). The results of the experiments in this paper are difficult to reconcile with either of these systems as currently understood.

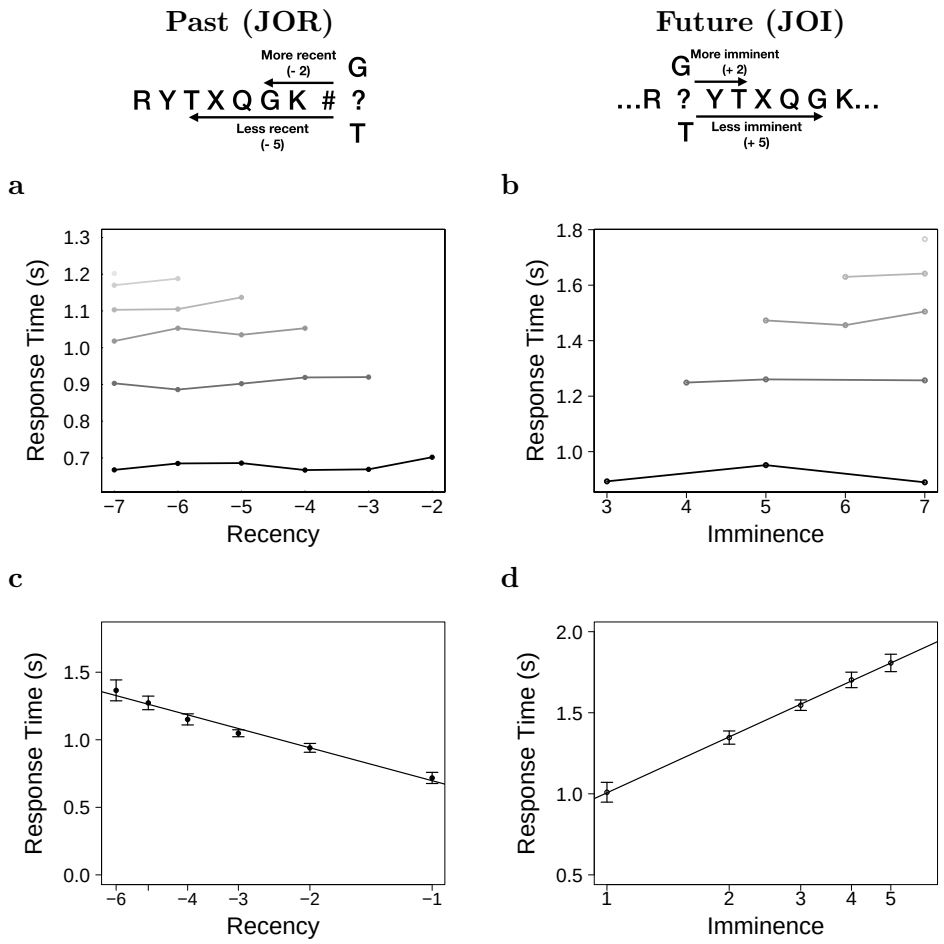


*Figure 2.* **RTs depend on the recency/imminence of the more recent/imminent probe but not on the recency/imminence of the less recent/imminent probe. a.** In the JOR task, median RTs for correct responses depend strongly on the recency of the more recent probe but not the recency of the less recent probe. **b.** Results from the JOI task show that the RT for correct responses depend strongly on the imminence of the more imminent probe. **c.** In JOR, median RT varies sub-linearly with recency. **d.** In JOI, the median RT varies sub-linearly with imminence of the more imminent item.

Model-free learning efficiently extracts predicted outcomes that follow from a stimulus. A major benefit model-free learning is that knowledge of outcomes are cached so that they can be rapidly accessed, avoiding the relatively costly computation associated with the model-based system. For instance, the successor representation (Dayan, 1993) is a computational method that enables rapid learning of exponentially-weighted future outcomes that follow a particular stimulus. The successor representation has many properties that have made it attractive to cognitive neuroscientists (e.g., Stachenfeld, Botvinick, & Gershman, 2016). Because the successor representation more strongly weights events that are likely to happen closer in time to the present, it would be expected to generate an effect of imminence. Moreover, this effect of imminence would be sublinear, as the successor representation weights future outcomes in an exponentially-discounted manner. This is at least roughly consistent with the results in Figure 2d. However, because both the strength of both probes are cached, there is no reason to expect that there would not be a strong interaction between the two probes and a robust distance effect. This is in sharp contrast to the lack of a distance effect on correct RT (horizontal lines in Fig. 2b) and the account of these results based on scanning.

In contrast a model-based account is perfectly consistent with a scanning model. The model-based system is believed to store the transition probabilities between stimuli (Daw et al., 2005). Predictions at successive time points can be simulated by repeatedly applying the matrix of transition probabilities to the current state. At each time step one could compare the predicted future to the probes and stop when a match is found. Although the model-based system is well positioned to build a scanning model, it is not clear why the growth in RTs would be sublinear. If one accepts the quantitative argument that the growth of RTs is logarithmic, that implies that the rate of applying the transition matrix must increase exponentially in order to account for the data.

The JOR results are consistent with the hypothesis that participants sequentially examine a logarithmically compressed representation of temporal history. This aligns with a large body of recent neurophysiological work on time cells in a variety of brain regions (MacDonald, Lepage, Eden, & Eichenbaum, 2011; Adler et al., 2012; Mello, Soares, & Paton, 2015; Tiganj, Cromer, Roy, Miller, & Howard, 2017). Time cells are sequentially activated neurons from which the recency of the stimulus that triggered the sequence can be reconstructed. They exhibit temporal compression (Salz et al., 2016; Howard et al., 2014; Kraus, Robinson, White, Eichenbaum, & Hasselmo, 2013), although it is not yet known if the compression is logarithmic. Logarithmically-compressed time cells are consistent with the predictions of computational models for constructing scale-invariant history (Shankar & Howard, 2013) and cognitive models built on that form of representation (Howard & Eichenbaum, 2013; Howard et al., 2015). In order to make sense of the results from the JOI experiment, one would have to develop a method to construct a logarithmically-compressed representation of the time at which future events take place. One possibility is that the future is constructed by translating the current state of the temporal history into the future (Shankar, Singh, & Howard, 2016). A method for constructing a representation of the future with the same properties as the scale-invariant representation of the past that has been hypothesized would be an important step towards building a model that makes sense of memory for the past and prediction of the future in the same framework (see also Gershman, Moore, Todd, Norman, & Sederberg, 2012; Gershman & Daw, 2017).

## Conclusion

This study introduced a novel Judgment of Imminence (JOI) task to study temporal order judgments for the future. The JOI task closely paralleled the design of the JOR task, allowing direct comparison. The classic finding in JOR is that the response time (RT) for correct judgments varies as a function of the distance to the more recent probe and does not depend on the distance to the less recent probe. This has suggested to many authors a backward self-terminating search model operating on a temporally-organized representation of the past. In the JOI task it was found that correct RT depends on the distance to the more imminent probe and does not depend on the distance to the less imminent probe. This suggests that a forward self-terminating search model operating on a temporally-organized representation of the future. This result supports the hypothesis that there is a deep symmetry between the psychological processes that underlie our ability to remember the past and make predictions about the future.

## Methods

*Experiment 1: Judgment of Recency (JOR)*

The procedure of this experiment follows the procedure of Experiment 2 of Hacker (1980) closely. Participants were presented with a list of 9, 11, or 13 consonants at the rate of 5.5 letters per second. At the end of the list, two of the last seven letters were chosen randomly and the participants were asked to indicate using left or right arrow key which of the two letters had appeared more recently. In Figure 1, G and T are presented as the probe items. Because G was presented more recently than T, the correct answer is G. In addition, participants were asked to respond with the up arrow key if they did not remember seeing either of the probe letters on the list. If the participant did not make a response within 6 s, the trial was terminated. Less than .004 of trials terminated without a response.

The distance to the more recent probe stimulus was varied from recency $-1$ (the last stimulus in the list) to $-6$. The recency of the less recent probe varied from $-2$ to $-7$. This leads to 21 possible combinations of recencies, which were presented in a random order. Each participant completed 320 trials.

There are several methodological differences between this procedure and the procedure of Experiment 2 of Hacker (1980). Unlike the Hacker (1980) study, in this experiment participants were never given foils that did not appear in the list. Also, in the Hacker (1980) study participants were not given the option to respond indicating that they did not remember either of the probes. The participants in the Hacker (1980) study were also more experienced in the task, experiencing a variety of presentation rates over several experimental sessions. In this experiment, participants received only one presentation rate in one session lasting about forty minutes.

*Participants.* The participants that participated in the study were drawn from the participant pool for Boston University's introductory psychology class. The study materials and protocol was approved by the Institutional Review Board at Boston University. 108 participants signed up for the study. One participant withdrew from the experiment. Data from 11 participants was excluded because their overall accuracy was no better than chance.

*Experiment 2: Judgment of Imminence (JOI)*

In order to probe the ability of participants to make relative order judgments of the future, participants were first trained on a probabilistic sequence. After training, the participants continued to see the same sequence interspersed with the probes consisting of two items that were from training. Participants indicated which of the two items would appear sooner (more imminent).

*Sequence generation*

For each participant, a unique transition ring similar to the one shown in Figure 1c was generated by randomly selecting 13 consonants. Throughout the duration of the experiment, the participants saw letters drawn from this ring. With a probability 0.9, the participants saw the next letter in the sequence. With probability 0.1, a random letter was chosen from the sequence that wasn't the letter just shown to the participant. A presentation list of 1550 letters was generated using this procedure and this was used presented to the participants one at a time over the course of the entire experiment from beginning to end. Each letter was displayed on the screen for 1.2 s followed by a fixation cross for .2 s.

*Training*

During the training round, participants were instructed to repeat each letter silently to themselves and that they would be tested on this sequence. After sixty letters were shown to the participants (about 1.5 mins), the initial training round ended and the experimenter initiated an abbreviated practice round for the actual task with the relative judgment probes (see below). The sequence continued for about two minutes. During this time participants were given twenty probes asking them to make predictions about relative order judgments of the future. The goal of this task practice round was to show the participants what the final task is so that they are motivated to pay attention to the sequence. After two minutes of this round, the training resumed and the participants were shown just the training sequence without any probes for another three minutes.

*Probes*

After each letter from the list, a probe was shown with a probability of 0.3. The probe consisted of two letters that were presented side by side. The participants were told that they will continue to see the letters one at a time and that sometimes two letters will be presented side by side. The task was to predict which of the two letters will appear sooner using the 'Left'/'Right' key. For instance, in Figure 1b, Q and K are presented and the participant is asked to predict which of the two is more imminent. The correct answer here is Q since it is three steps from the current position in the sequence *vs* K which is five steps in the sequence. The instructions clarified that they were not necessarily required to choose the next letter in the sequence (although that would sometimes be the case) but the one that would appear sooner. Probes were not shown one after another; at least one letter from the sequence was displayed before the next probe. If the participants failed to make a response in 4 s, they were shown a message,"Faster" for 0.2 s and then the task continued. About .013 of trials resulted in a time out. On average the participants were shown 406 probes. The first 10% of the probes (these included the probes shown in training) were

omitted from further analyses. The post-training round lasted for about thirty minutes with two intermediate breaks.

Based on pilot studies and debriefing interviews, it was found that participants were particularly cognizant of pairwise associations between the letters in the sequence. In order to eliminate an account based on simple pairwise associations between the probes, we did not present probes that contained stimuli that would be expected to be presented successively. For instance, in Figure 1b, if the more imminent item in the probe was X, the less imminent item could not be Q. The two probe letters were sampled from the next 7 positions in the transition ring. The imminence of the more imminent item varied from 1 (next in the sequence) to 5. For each of these more imminent items, the less imminent items were chosen using pre-determined combinations. The points were sampled such that where possible three less imminent probes for each more imminent probe. For instance, when the imminence of the more imminent probe was 1 (next item), the less imminent item was chosen from imminencies of 3, 5, or 7. When the more imminent probe was at an imminence of 3, the less imminent probe was chosen from imminencies of 5, 6, or 7. Since only the next 7 positions are probed, a more imminent probe at an imminence of 4 can only be paired with a less imminent probe at an imminence of 6 and 7. A more imminent probe at an imminence of 6 can only be probed with a more imminent pairing at an imminence of 7.

*Participants*

Sixty healthy young adults were recruited from Boston University to participate in one session each and were paid $15 per hour for their time. The study materials and protocol were approved by the Institutional Review Board at Boston University. Three participants performed no better than chance; their data was excluded from further analysis.

*Strategy for analyzing the two recency/imminence variables*

On each trial the participant is given two probe stimuli. We are interested to know if the recency/imminence of both of the probes affects correct RT or if only the recency/imminence of the correct probe does. One might have simply put these two variables (perhaps recoding them using a distance between the two probes) into a linear regression. However the recencies/imminencies are not sampled equally. Moreover the effect of recency/imminence appears to be non-linear (Fig. 2c,d). Using a straightforward linear model carries the risk of a spurious distance effect due to residuals from a non-linear and unequally sampled relationship.

In order to control for these effects, the following two step strategy was used. First the distance effect was quantified by calculating the slopes of the lines joining the dependent variable as a function of the less salient independent variable. For instance, in Figure 2a, the lines for the correct RT appear to be flat. Thus the median RT does not appear to depend on the recency of the less recent probe. In order to quantify this intuition, the slope of each line was calculated for each participant. In order to ascertain whether these slopes are meaningfully different from 0 or not, a Bayesian t-test (Rouder et al., 2009) was performed on the slopes obtained for each participant. If a particular variable did not contribute to the dependent variable, it was excluded as a factor in a subsequent linear mixed effect analysis. Thus the purpose of this analysis was simply to ascertain whether a factor meaningfully contributed to the dependent variable. Once the relevant factors affecting the dependent

variable were ascertained, further analyses were performed with the relevant variables. The function of the linear mixed effect analysis is simply to determine whether the apparent effect of the more salient variable is statistically reliable.

## References


Addis, D. R., Wong, A. T., & Schacter, D. L. (2007). Remembering the past and imagining the future: Common and distinct neural substrates during event construction and elaboration. *Neuropsychologia*, *45*(7), 1363–1377.

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

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

Bar, M. (2009). The proactive brain: memory for predictions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1235-43. doi: 10.1098/rstb.2008.0310

Bialek, W. (2012). *Biophysics: searching for principles.* Princeton University Press.

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

Daw, N. D., & Dayan, P. (2014). The algorithmic anatomy of model-based evaluation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1655). doi: 10.1098/rstb.2013.0478

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature neuroscience*, *8*(12), 1704–1711.

Dayan, P. (1993). Improving generalization for temporal difference learning: The successor representation. *Neural Computation*, *5*(4), 613–624.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, *11*, 127-138.

Gershman, S. J., & Daw, N. D. (2017). Reinforcement learning and episodic memory in humans and animals: An integrative framework. *Annual Review of Psychology*, *68*, 101–128.

Gershman, S. J., Moore, C. D., Todd, M. T., Norman, K. A., & Sederberg, P. B. (2012). The successor representation and temporal context. *Neural Computation*, *24*(6), 1553–1568.

Hacker, M. J. (1980). Speed and accuracy of recency judgments for events in short-term memory. *Journal of Experimental Psychology: Human Learning and Memory*, *15*, 846-858.

Hassabis, D., Kumaran, D., Vann, S. D., & Maguire, E. A. (2007). Patients with hippocampal amnesia cannot imagine new experiences. *Proceedings of the National Academy of Sciences USA*, *104*(5), 1726-31. doi: 10.1073/pnas.0610561104

Hockley, W. E. (1984). Analysis of response time distributions in the study of cognitive processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(4), 598-615.

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

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

James, W. (1890). *The principles of psychology.* New York: Holt.

Jeffreys, H. (1998). *The theory of probability.* OUP Oxford.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*, 773-795.

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

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.

McElree, B., & Dosher, B. A. (1993). Serial recovery processes in the recovery of order information. *Journal of Experimental Psychology: General*, *122*, 291-315.

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

Momennejad, I., Russek, E. M., Cheong, J. H., Botvinick, M. M., Daw, N., & Gershman, S. J. (2017). The successor representation in human reinforcement learning. *Nature Human Behaviour*, *1*(9), 680.

Murdock, B. B. (1974). *Human memory: Theory and data*. Potomac, MD: Erlbaum.

Muter, P. (1979). Response latencies in discriminations of recency. *Journal of Experimental Psychology: Human Learning and Memory*, *5*, 160-169.

Palmer, S. E., Marre, O., Berry, M. J., 2nd, & Bialek, W. (2015). Predictive information in a sensory population. *Proceedings of the National Academy of Sciences USA*, *112*(22), 6908-13. doi: 10.1073/pnas.1506855112

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237.

Russek, E. M., Momennejad, I., Botvinick, M. M., Gershman, S. J., & Daw, N. D. (2017). Predictive representations can link model-based reinforcement learning to model-free mechanisms. *bioRxiv*, 083857.

Salz, D. M., Tiganj, Z., Khasnabish, S., Kohley, A., Sheehan, D., Howard, M. W., & Eichenbaum, H. (2016). Time cells in hippocampal area CA3. *Journal of Neuroscience*, *36*, 7476-7484.

Schacter, D. L., Addis, D. R., & Buckner, R. L. (2007). Remembering the past to imagine the future: the prospective brain. *Nature Reviews, Neuroscience*, *8*(9), 657-661.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*, 1593-1599.

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

Shankar, K. H., Singh, I., & Howard, M. W. (2016). Neural mechanism to simulate a scale-invariant future. *Neural Computation*, *28*, 2594–2627.

Stachenfeld, K. L., Botvinick, M. M., & Gershman, S. J. (2016). The hippocampus as a predictive map. *bioRxiv*, 097170.

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

Tulving, E. (1985). How many memory systems are there? *American Psychologist*, *40*, 385-398.

Wetzels, R., & Wagenmakers, E.-J. (2012). A default Bayesian hypothesis test for correlations and partial correlations. *Psychonomic Bulletin & Review*, *19*(6), 1057–1064.

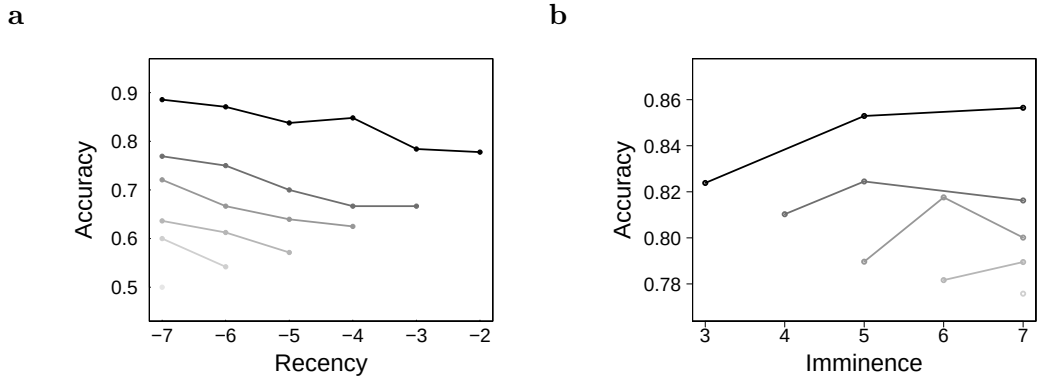


*Figure S1.* Accuracy in the JOR task and the JOI task. **a.** Accuracy in the JOR task showed a reliable recency effect and a reliable distance effect. **b.** Accuracy in the JOI task was higher for probes that were closer in time to the present but did not show a reliable distance effect. See text for details.

## Supplemental Information

*Supplementary JOR results*

*JOR accuracy showed a recency effect and a distance effect.* The probability that participants selected the more recent probe was $.70 \pm 0.01$. The accuracy was $.82 \pm .01$ when the recency of the more recent probe was $-1$ and dropped to $.49 \pm .02$ when the recency was $-6$. At recency $-6$ the probability of choosing the more recent probe was not different from chance (Chi-squared prop test, $\chi^2(96) = 89.1$, p-value not significant). Recency $-5$ had an accuracy of $0.56 \pm 0.1$ and was significantly higher than chance (Chi-squared prop test, $\chi^2(96) = 142.6$, $p < 0.01$).

Accuracy also depended on the temporal distance between the more recent probe and the less recent probe. For a recency of the more recent probe, the accuracy improved as the recency of the less recent probe decreased (distance effect). The upward-sloping lines in Figure 2a indicate the presence of this distance effect. To quantify the distance effect for each participant, we calculated the slope of each line in Figure 2a. A Bayesian t-test (Rouder et al., 2009) on the obtained slopes revealed "decisive evidence" (Wetzels & Wagenmakers, 2012; Kass & Raftery, 1995; Jeffreys, 1998) favoring the hypothesis that the slopes are different from 0 (JZS Bayes Factor $> 10^2$).

The effects of recencies of the two probes on accuracy was quantified using a linear mixed effect analysis with independent intercepts for each participant. The accuracy increased with an increase in the recency of the more recent probe by $.078 \pm .002$, $t(1918) = -31.9$, $p < 0.01$ per unit change in recency. Accuracy also increased with the recency of the less recent probe by $.023 \pm .002$, $t(1918) = 9.73$, $p < 0.01$ per unit. These findings are consistent with the findings from prior studies.

*Incorrect RT depended only on the recency of the selected probe in the JOR task.* In a self-terminating backward scanning model, if the scan misses the more recent probe, it would then terminate on the less recent probe. These responses would be errors and the scanning time for these errors would depend on the recency to the less recent probe.

Given the overall error rate of $.30 \pm 0.01$, there are less than half the number of observations for incorrect RTs as there are for correct RTs. Also note that the number of errors is not evenly distributed over recencies, so that some points have many fewer observations than others. Nonetheless, error RTs appear to depend reliably on the recency to the less recent item. There does not appear to be a strong effect of the recency to the more recent probe.

To evaluate whether there was an effect of the recency to the more recent probe we calculated the slope of the distance effect for each value of the recency to the less recent probe. This is analogous to the distance effect calculation for correct RTs except the distance effect is calculated separately for the recency to the less recent probe rather than the more recent probe. That is, for errors we computed a slope for each cluster of points in Figure 2c rather than across each line. A Bayesian t-test showed "strong evidence" (Wetzels & Wagenmakers, 2012; Kass & Raftery, 1995; Jeffreys, 1998) favoring the hypothesis that the slopes of the median RTs as a function of the more recent recency are not different from 0 (JZS Bayes Factor = 14.5).

A linear mixed effects analysis, allowing each participant to have an independent intercept, and the less recent recency as regressor showed a significant effect of the recency to the less recent probe on the median RT of incorrect responses, $.033 \pm .007$ s, $t(473) = 4.9$, $p < 0.001$.

*Supplementary JOI results*

*JOI accuracy depended only on the imminence of the more imminent probe.* The overall accuracy in this task was $.81 \pm .02$. The accuracy varied from $.84 \pm .02$ when the most imminent imminency was 1 to $.77 \pm .02$ when the most imminent imminency was 5. In order to check if the somewhat upward-sloping lines in Figure 2b indicate the presence of a reliable distance effect, the slopes of each of the lines was calculated for every participant. A Bayesian t-test (Rouder et al., 2009) on the obtained slopes revealed "substantial evidence" (Wetzels & Wagenmakers, 2012; Kass & Raftery, 1995; Jeffreys, 1998) favoring the hypothesis that the slopes are not different from 0 (JZS Bayes Factor 6.5).

Thus the accuracy depended on the imminence to the more imminent probe alone. The effects of the more imminent probe on accuracy was quantified using a linear mixed analysis with independent intercepts for each participant. Accuracy decreased with an increase in the imminence to the more imminent probe by $.017 \pm .002$, $t(227) = -6.9$, $p < 0.01$ per unit change in imminence. The finding that participants were more accurate based on the imminence of the more imminent probe is different from the accuracy results seen in JOR tasks (Hacker, 1980) where accuracy depends on both the more recent and the less recent probes. The overall accuracy is the JOR task is lower than the accuracy seen in this task and drops close to chance for the longest recency unlike this task where the accuracy for the longest imminence is reliably above chance.

*Incorrect RT showed mixed results in the JOI task.* When participants make errors, the forward scanning model predicts that the RT should depend on the imminence of the less imminent probe. However this was not observed in the error responses here, perhaps because accuracy was much higher than the JOR task. The overall error rate was $.19 \pm .02$

and for each participant on average 6 error responses per condition were obtained. To evaluate whether the more imminent probe contributed to the change in median RTs of a particular less imminent probe (distance effect), the slope across the more imminent imminencies for each less imminent imminency was calculated . This is analogous to the distance effect calculation for correct RTs except the distance effect is calculated separately for the imminence of the less imminent probe rather than the more imminent probe. The less imminent imminencies of 3 and 4 only have one possible combination so a slope could not be calculated for these imminencies. Across the remaining slopes, a Bayesian t-test showed that the evidence for a distance effect was "barely worth a mention" (JZS Bayes Factor = 1.1) (Wetzels & Wagenmakers, 2012; Kass & Raftery, 1995; Jeffreys, 1998).

Given the overall error rate of .19, there are less than one fourth the number of observations for incorrect RTs as there are for correct RTs. While a linear mixed effects analysis, allowing each participant to have an independent intercept, and both imminencies as regressors, showed a significant effect of the imminence of the more imminent probe on the median RT of incorrect responses, $.04 \pm .01$ s, $t(586) = 3.0$, $p < 0.001$, a linear model with both the more and the less imminent imminencies did not show any significant effect of the more or the less imminent probe, $.04 \pm .02$ s, $t(641) = 1.8$ and $.01 \pm .01$ s, $t(641) = .4$ respectively.