Time-Conjunctive Representations Of Future Events

Stuart W. Babcock[1], Marc W. Howard[2], and Joseph T. McGuire[2]

[1]Graduate Program for Neuroscience, Boston University

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

Abstract

It is widely accepted that people can predict the relative imminence of future events. However, it is unknown whether the timing of future events is represented using only a "strength-like" estimate or if future events are represented conjunctively with their position on a mental timeline. We examined how people judge temporal relationships among anticipated future events using the novel Judgment of Anticipated Co-Occurence (JACO) task. Participants were initially trained on a stream of letters sampled from a probabilistically repeating sequence. During test trials, the stream was interrupted with pairs of probe letters and the participants' task was to choose the probe letter they expected to appear in the stream during a lagged target window 4-6 items (4.3-8.5 seconds) in the future. Participants performed above chance as they gained experience with the task. Because the correct item was sometimes the more imminent probe letter and other times the less imminent probe letter, these results rule out the possibility that participants relied solely on thresholding a strength-like estimate of temporal imminence. Rather, these results suggest that participants held 1) temporally organized predictions of the future letters in the stream, 2) a temporal estimate of the lagged target window, and 3) some means to compare the two and evaluate their temporal alignment. Response time increased with the lag to the more imminent probe letter, suggesting that participants accessed the future sequentially in a manner that mirrors scanning processes previously proposed to operate on memory representations in the short-term judgment of recency task.

Introduction

"It is perfectly true, as the philosophers say, that life must be understood backwards. But they forget the other proposition, that it must be lived forwards." - *Søren Kierkegaard* (Kierkegaard, 1938)

The capacity for prospection is the guiding force of human action. By accurately representing the future, we can imagine how events will unfold and can make decisions accordingly in the present: if I want to become a doctor in the future, I must study physiology now; if I want to be healthy in my old age, I must exercise and eat properly now; if I want to send my children to university, I must save for that expense now. This is true of shorter time scales as well: moment-to-moment predictions of upcoming stimuli help guide perception, action, and learning, and may be a fundamental aspect of cognitive function (Clark, 2013).

Prospection, short- or long-term, has been argued to rely on some of the same cognitive processes that support memory (Tulving, 1985a). A number of studies have demonstrated that the neural bases of these processes overlap (Atance & O'Neill, 2001; Addis, Wong, & Schacter, 2007; Schacter, Addis, & Buckner, 2007; Szpunar, Watson, & McDermott, 2007; Bar, 2009; Spreng, Mar, & Kim, 2009), suggesting that simulation of an event, whether experienced (past) or imagined (future), relies on brain structures such as the medial temporal lobe (Okuda et al., 2003) and ventromedial prefrontal cortex (Hassabis & Maguire, 2007). Prospection is impaired in amnesia (Tulving, 1985b; Klein, Loftus, & Kihlstrom, 2002; Kwan, Carson, Addis, & Rosenbaum, 2010), the level of detail in prospective and remembered experiences is correlated across individuals (Williams et al., 1996), and prospective simulations are more detailed when more relevant memory details are available (Szpunar & McDermott, 2008).

An outstanding question is whether time serves the same organizing role for prospection as it does for memory. Episodic memory prominently includes information about the time at which events occurred: the accuracy and latency of memory retrieval depends on recency (Shepard & Teghtsoonian, 1961; Monsell, 1978; Glenberg et al., 1980; Hockley, 1982; Donkin

& Nosofsky, 2012) and temporal contiguity effects in free recall imply that remembered events are bound to a temporal context, which is retrieved with an event and serves as a cue for adjacent events (Murdock, 1962; Murdock & Okada, 1970; Murdock & Metcalfe, 1978; Kahana, 1996). Response times in the short-term judgment of recency task, in which participants judge which of two probe items occurred more recently in a stimulus stream, are consistent with a self-terminating backward scan along a temporally ordered memory representation (Muter, 1979; Hacker, 1980; Hockley, 1984; McElree & Dosher, 1993). These findings collectively imply a temporal axis along which events in memory are ordered and accessed.

There is some evidence already that future events, like past events, are represented along a temporal axis. Imagined future events are weaker and less detailed if they are temporally more remote, like a mirrored future-analogue of forgetting (D'Argembeau & Van der Linden, 2004). Rewarding events are discounted according to their anticipated delay (Samuelson, 1937; Frederick, Loewenstein, & O'Donoghue, 2002), and episodic prospection has incidental effects on temporal discounting (Peters & Büchel, 2010; Daniel, Stanton, & Epstein, 2013; Liu, Feng, Chen, & Li, 2013). At a neural level, theoretical work on temporal processing predicts the existence of temporally ordered future representations (Shankar, Singh, & Howard, 2016; Tiganj, Gershman, Sederberg, & Howard, 2019).

In the present investigation, we tested the temporal organization of future event representations by asking participants to judge temporal relationships among independent anticipated events in the near future. This form of near-future prospection is analogous to the near-past memory utilized in the judgment of recency task. To establish a predictable future, we created a series of stimuli that followed a regular sequence, analogous to a schematic script (Abelson, 1981) or the sequence learning that occurs in serial reaction time tasks (Willingham, Nissen, & Bullemer, 1989). Other recent work using a similar sequentially organized environment has demonstrated that people can successfully judge which of two stimuli is more imminent, a future-oriented analogue to judgments of recency (Singh & Howard, 2017).

Here, we tested participants' ability to make more complex judgments about which events would coincide with a specific time window in the future.

Our experiment was designed to assess two alternative ways in which the timing of future events could be cognitively represented. One possibility is that an event's temporal proximity is represented by a strength-like scalar. Strength-like representations of the timing of events have a role in a number of models of memory and reward anticipation (Hinrichs, 1970; Wickelgren, 1970; Sutton & Barto, 1981; Schultz, Dayan, & Montague, 1997), although such strength models have been unable to explain important empirical results in the memory domain (Hintzman & Block, 1971; Hintzman, Block, & Summers, 1973; Hintzman, 2010). In our task, merely thresholding a scalar, strength-like representation of an event's temporal proximity would not support successful performance, because the task was equally likely to require selecting the more- or less-imminent item. Successful performance in the present task would suggest, alternatively, that future events can be cognitively represented in conjunction with their position on a prospective mental timeline. Such a temporally organized representation of future events could permit sequential scanning along the timeline, enabling participants to imagine the future as it would unfold through time and judge which future events are expected to co-occur.
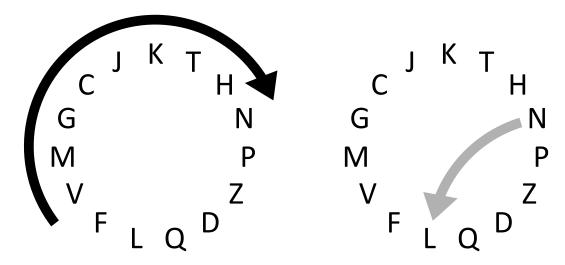
Methods

Participants were recruited using Boston University's student jobs website. The advertisement informed participants that they would complete a one-hour cognitive task on a computer and would be compensated $12 for their time. Once in the lab, participants were informed that they would receive additional compensation based on their performance on the task ($0.15 per correct response, $6.00-$12.75 per participant). 13 participants (aged 18-25, 9 female) were recruited as an initial sample. We proceeded to recruit a second sample of 13 participants (aged 18-28, 10 female) as a replication test of the results from the

first 13. From this second data set, one participant's data file was mistakenly written to a non-existent directory, losing the data and leaving us with 12 participants in our replication sample.

Participants completed the task on a Macbook Pro running OS X El Capitan (version 10.11.6). The experiment was written in Python using PsychoPy2 version 1.85.1 (Peirce, 2007). Participants completed the task individually in a testing room under the supervision of the primary author.

The Judgment of Anticipated Co-Occurence (JACO) task required participants to make predictions about the future temporal overlap of 2 predictable events, a probe letter and a lagged target window. Participants were told that they would view a stream of letters, consisting of a repeating sequence of 15 letters which usually progressed in a fixed order but occasionally jumped to a random point in the sequence before continuing in order (Fig. 1). Letters were viewed one at a time for 1.2 seconds with a 0.2 second blank inter-item interval. Participants completed 4 task blocks (1 training block and 3 testing blocks) which lasted approximately 12 minutes each.

Usually, the stream repeated in order... Sometimes, the stream jumped...

Figure 1: The stream proceeded in a patterned way. Usually, the stream proceeded in the order of the sequence. Shortly before each trial, the stream would jump to an unpredictable point in the sequence. After the jump, the stream proceeded in the order of the sequence until the trial was over.

In the training phase (Block 1), participants first viewed the stream of letters for 4 minutes during which they were instructed to learn the sequence. After these 4 minutes, the stream was occasionally interrupted by a 2-alternative forced-choice trial that prompted the participant to predict the next letter in the stream. The visual prompt showed an empty purple square in the center of the screen where the stream letter normally appeared, flanked by 2 probe letters. One of these 2 probe letters was the next letter in the learned sequence (the target) and the other letter was not (the lure). Participants had up to 4 seconds to respond by pressing the left or right arrow key. After the response, the chosen probe letter and the purple square remained on the screen while the next letter of the stream was presented. If the next letter matched the chosen probe letter, participants heard a "ding," signifying that $0.15 had been added to their earnings. Block 1 had 15 trials, and each letter in the sequence occurred once as a target and once as a lure.

In the testing phase (Blocks 2, 3, and 4), the participants' task was to judge which letter would appear during a lagged target window, which encompassed the letters at lags 4, 5, and 6 in the future (Fig. 2). Before the testing phase, participants saw a short instructional example that showed the behavior of the lagged target window. The visual prompt for trials showed an empty yellow square in the center of the screen flanked by 2 probe letters; a target, which would occur within the lagged target window, and a lure, which would occur within 2 letters before or after the window (lags 2, 3, 7, or 8). As in the training phase, participants had up to 4 seconds to respond. After the response, the chosen probe letter remained on the screen but the yellow square disappeared as the stream progressed through lags 1, 2, and 3. After lag 3, the yellow square reappeared and remained through lags 4, 5, and 6. After lag 6, the chosen probe letter and the yellow square disappeared for good. If the chosen probe letter matched a letter that appeared in the stream while the yellow square was visible, participants heard a "ding," signifying that $0.15 had been added to their earnings.

In the testing phase, the stream had the same sequential structure as it did in the training phase (Fig. 1). Eight letters after each trial, the sequence jumped to a point 4 to 7
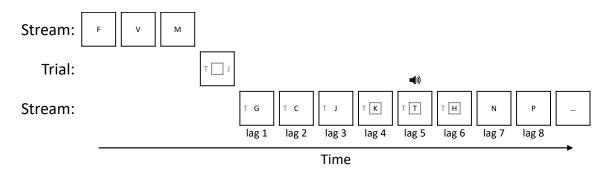
Figure 2: A sequence of screens showing what the participant would see on a correctly answered trial in the testing phase. The auditory cue occurred only if the stream letter matched the chosen probe letter while the square was on the screen; the cue represented earnings of $0.15.

letters before where the next trial would appear. The jumps served to decouple recent past experience from future predictions, so responses could not be reliably guided by the relative recency of the 2 probe items (Fig. 3). Participants were invited to take short breaks between blocks.
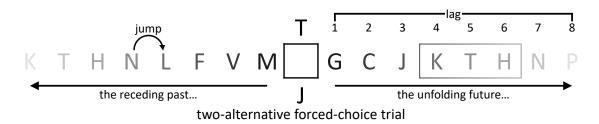

Figure 3: A hypothesized internal representation during the testing phase. When a trial occurred, the participant judged which of the 2 probe letters would occur in the lagged target window, using the remembered past to orient placement in the sequence and a simulated future to access relevant representations of future events.

For each pair of probe letters, a target at lag 4, 5, or 6 positions in the future was paired with a lure at lag 2, 3, 7, or 8 positions in the future with the constraint that the target and lure were not adjacent in the sequence. Probe letters were limited to forward lags 1-8 so that each probe letter's forward-looking lag was shorter than or equal to its backward-looking lag within the learned sequence (for example, forward lag 7 is equivalent to backward lag 9 in our 15-letter sequence). Given these criteria, 45 unique probe letter pairings were possible. From these pairings, we created 90 trials such that each unique pair was presented twice, and each probe letter within the pair was a target once and a lure once. This trial-construction scheme

ensured that each sequence letter, when presented as a probe letter, was equally likely to be a target or lure and equally likely to be the more or less imminent of the 2 probe letters. We further arranged a relatively uniform sampling of all possible target-lure lag combinations. The 90 trials were randomly shuffled, and each of Blocks 2, 3, and 4 contained 30 trials.

All analyses were written in R (version 3.5.1) using RStudio (version 1.1.453). All figures were created using the same software. Packages were used to calculate effect sizes (Navarro, 2015), do power analyses (Champely, 2018), draw scatterplots (Eklund, 2016), create mixed-effects regressions (Bates, Maechler, Bolker, & Walker, 2015), plot confidence intervals (Lemon, 2006), execute variance and closeness tests (Merkle & You, 2018; Wang & Merkle, 2018), and make hierarchical Bayesian models (Bürkner, 2017; Carpenter et al., 2017; Bürkner, 2018; Stan Development Team, 2018).

Results

The response on each trial was categorized as correct, incorrect, or missed (no response). All analyses presented hereafter consider only trials with responses. Missed trials (18.1% in training phase; 2.31% in testing phase) were excluded from subsequent analyses.

Training-phase accuracy

Training phase performance (Block 1) was above chance, indicating that participants learned the sequence. Group accuracy was significantly above chance ($p < 0.001$, 2-sided Wilcoxon rank-sum test) with a median accuracy of 71.4% ($\sigma = 0.155$; Fig. 4a). The effect of trial number on accuracy was not significant ($p = 0.155$, mixed-effects probit regression), but by the end of the block the estimated mean accuracy was 84.0% (mixed-effects probit regression; Fig. 4b). Mean response time was 1.35 seconds ($\sigma = 0.687$).
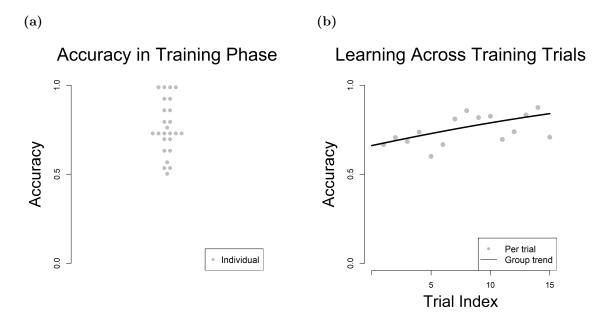
(a)   **(b)**

Figure 4: Participants learned the sequence during training (Block 1). **(a)** As a group, participants performed significantly above chance on training trials ($p < 0.001$, 2-sided Wilcoxon rank-sum test). The y-axis shows accuracy. Median participant accuracy was 71.4% ($\sigma = 0.155$). **(b)** Accuracy over the course of the block. The x-axis shows the trial number and the y-axis shows accuracy. By the end of training, estimated mean accuracy was 84.0% (mixed-effects probit regression).

Testing-phase accuracy

Accuracy increased over the course of the testing phase (Blocks 2, 3, and 4), indicating that participants learned how to complete the task. Similar trajectories were observed in both samples (Table 1).

Group-Level Mean Accuracy Across Testing Blocks

	Block 2	Block 3	Block 4
Sample 1	0.556 ($\sigma = 0.142$)	0.575 ($\sigma = 0.147$)	0.655 ($\sigma = 0.142$)
Sample 2	0.523 ($\sigma = 0.0991$)	0.560 ($\sigma = 0.0933$)	0.684 ($\sigma = 0.0971$)
Combined Sample	0.540 ($\sigma = 0.122$)	0.568 ($\sigma = 0.122$)	0.669 ($\sigma = 0.121$)

Table 1: Accuracy rates by testing block for each sample. Accuracy increased from Block 2 through Block 4 for both samples.

In Block 4, group accuracy was significantly above chance in both samples ($p = 0.004$ and $p = 0.004$, 2-sided Wilcoxon rank-sum test) and in the combined sample ($p < 0.001$, 2-sided Wilcoxon rank-sum test; Fig. 5a). Effect size was large in both samples (Cohen's $d = 1.10$ and Cohen's $d = 1.89$) and the combined sample (Cohen's $d = 1.40$). 4 of 13 participants

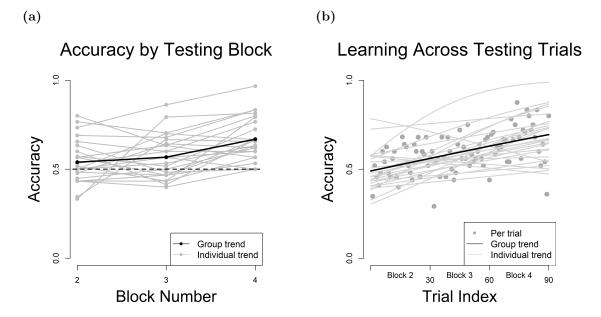**(a)**   **(b)**

Figure 5: **(a)** Accuracy increased across testing blocks. The x-axis shows the testing block number (2-4) and the y-axis shows accuracy for the trials in that block. Participant- and group-level trajectories are shown. The dashed line shows chance-level accuracy. Group-level performance significantly exceeded chance in Block 3 ($p = 0.017$, 2-sided Wilcoxon rank-sum test) and Block 4 ($p < 0.001$, 2-sided Wilcoxon rank-sum test), but not in Block 2 ($p = 0.169$, 2-sided Wilcoxon rank-sum test). Individual participants significantly exceeded chance as well: 3 in Block 2, 3 in Block 3, and 8 in Block 4 ($>70\%$ accuracy, $p < 0.05$, 2-sided exact binomial test). No participants performed below chance in the final block. **(b)** Participants learned continuously over the course of the experiment. The x-axis shows the trial number and the y-axis shows accuracy. Points show the group-level accuracy for each trial number. Grey lines show probit-regression estimates of individual-level accuracy across trials and the black line shows a mixed-effects probit-regression estimate of group-level accuracy across trials. There was a clear effect of trial number on group-level accuracy ($p < 0.001$, mixed-effects probit regression) which was also apparent in most individuals.

(Sample 1) and 4 of 12 participants (Sample 2) performed significantly above chance in individual-level significance tests ($>70\%$ accuracy, $p < 0.05$, 2-sided exact binomial test). No participants performed below chance in Block 4. Due to the performance similarities between the two samples, subsequent analyses consider only the combined sample.

Participants exhibited progressively increasing accuracy across trials, and there was a significant effect of trial number on accuracy ($p < 0.001$, mixed-effects probit regression; Fig. 5b). Estimated mean accuracy ranged from 49.3% at the start of the testing phase to 69.5% at the end of the testing phase (mixed-effects probit regression). The group-level probit fit can be interpreted in terms of a linear change in the sensitivity index (d') as a function of trial number (DeCarlo, 2012): the estimated sensitivity index ranged from -0.025

at the start of the testing phase to 0.720 at the end of the testing phase. Comparing accuracy rates from the training block (Block 1) and the last testing block (Block 4), we found a weak positive correlation ($\rho = 0.186$, $p = 0.385$, Spearman), indicating that we could neither strongly confirm nor rule out an association between training success and task performance.

Response time and accuracy by lag

We tested the hypothesis that future-event representations were accessed via a self-terminating forward scan along a temporally ordered prospective representation of the stimulus sequence, akin to scanning processes that have been proposed to operate in short-term judgment of recency tasks (Muter, 1979; Hacker, 1980; Hockley, 1984; McElree & Dosher, 1993). The scanning process is assumed to proceed forward from the present time, serially examining each anticipated item and terminating when either the target or lure is encountered, whichever comes first. The scanning hypothesis predicts an increase in response time as a function of the lag to the more imminent of the two probe items. Previous studies of retrospective scanning processes suggest that the increase in response time as a function of lag is sublinear, indicating that scanning takes place along a compressed timeline of events (Brown, Neath, & Chater, 2007; Howard, Shankar, Aue, & Criss, 2015; Singh & Howard, 2017).

To focus on the mechanisms of successful task performance, analyses of response times were restricted to correct trials in Block 4. The analyses were carried out both for the full sample (Figs. 6a-b) and a subset of high-performing individuals (>66.7% accuracy, $n = 8$; Figs. 6c-d), based on the assumption that error rates imply a similar number of trials could have been correct by chance.

Response times increased as a function of the lag to the more imminent probe letter (Figs. 6a, 6c). This effect was consistent with the hypothesized self-terminating forward scan along a temporally ordered prospective representation of the stimulus sequence. The effects of lag and log lag on response time were significant for the full sample and the high-performing
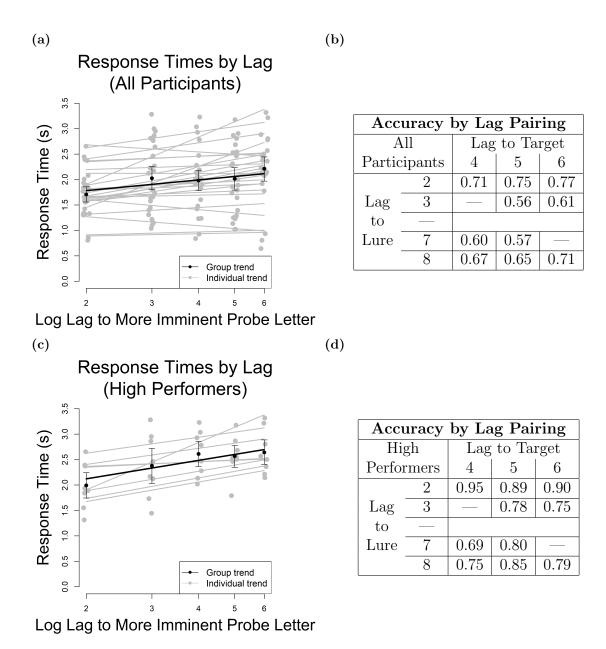
(a)

Response Times by Lag
(All Participants)

Response Time (s) vs *Log Lag to More Imminent Probe Letter*

Group trend
Individual trend

(b)

Accuracy by Lag Pairing			
All	Lag to Target		
Participants	4	5	6
2	0.71	0.75	0.77
Lag 3	—	0.56	0.61
to			
Lure 7	0.60	0.57	—
8	0.67	0.65	0.71

(c)

Response Times by Lag
(High Performers)

Response Time (s) vs *Log Lag to More Imminent Probe Letter*

Group trend
Individual trend

(d)

Accuracy by Lag Pairing			
High	Lag to Target		
Performers	4	5	6
2	0.95	0.89	0.90
Lag 3	—	0.78	0.75
to			
Lure 7	0.69	0.80	—
8	0.75	0.85	0.79

Figure 6: **(a)** Response time increased with the lag to the more imminent probe letter. The x-axis shows the log lag to the more imminent probe letter and the y-axis shows response time. Data are from final testing block (Block 4) trials which were correctly answered ($n = 25$). Median response times are shown for individuals and means of medians are used for the group. Confidence intervals (95%) were calculated using bootstrap of the mean. The increase in response time with respect to the lag to the more imminent item was consistent with a self-terminating forward scan. **(b)** Accuracy was influenced by whether lures were further from the window (lags 2 and 8) or nearer (lags 3 and 7; $p = 0.003$, mixed-effects probit regression). Data are from final testing block (Block 4) trials ($n = 25$). **(c, d)** Equivalent results were observed in analyses restricted to high-performing participants ($n = 8$), both for response times **(c)** and accuracy **(d)**.

subset ($p < 0.001$, mixed-effects linear regression). As the linear and logarithmic models were

non-nested, we used Vuong's method to determine whether the models were distinguishable

and, if so, which one more closely fit the data (Vuong, 1989). The difference in fit between the linear and logarithmic models was not significant, either in the full sample ($p = 0.311$, Vuong's closeness test) or in the high-performing subset ($p = 0.238$, Vuong's closeness test), although the models were distinguishable in both cases ($p = 0.018$ & $p = 0.011$, Vuong's variance test).

Accuracy was influenced by whether the lure was nearer to the target window (lure lag 3 or 7) or further from it (lure lag 2 or 8). We analyzed accuracy as a function of three categorical factors: inner/outer lure position (lag 3 or 7 vs. lag 2 or 8), early/late lure position (lag 2 or 3 vs. lag 7 or 8), and inner/outer target position (lag 5 vs. lag 4 or 6). We observed a significant effect on accuracy of inner/outer lure position ($p = 0.003$) and no effect of early/late lure position ($p = 0.165$) or inner/outer target position ($p = 0.786$; Fig. 6b). The high-performing subset ($n = 8$) showed a similar pattern, with a significant effect on accuracy of inner/outer lure position ($p = 0.050$) and no significant effects of the other factors (Fig. 6d).

Testing alternative explanations

Finally, we considered the alternative possibility that accuracy increased across trials due to some form of stimulus-based reinforcement learning that did not require accessing a representation of the future. If a simple form of reinforcement learning were a viable strategy, then participants might have been able to attain above-chance accuracy without relying on a time-conjunctive representation of future events.

The task was designed to rule out a simple reinforcement learning strategy operating on individual probe letters. Each letter appeared as a probe item a total of 12 times: 6 times as a target and 6 times as a lure. Therefore, a strategy of preferentially selecting previously rewarded probe letters would not yield above-chance performance.

Another potential reinforcement learning strategy would act on sets of paired probe letters. Each unique pair of probe letters was presented twice during the experiment. The

2 presentations occurred at different points in the sequence such that each of the 2 probe letters in the pair was a target once and a lure once. Therefore, it is possible, if implausible, that if participants remembered which letter was the target during the first presentation of a pair, they could perfectly predict that the other letter would be the target in the second presentation of the same pair.

To evaluate this possibility, we examined accuracy rates on first and second presentations of a given probe letter pair separately as a function of trial index (Fig. 7). We did not observe any significant difference in performance between the 2 presentations, neither in the intercept ($p = 0.815$, mixed-effects probit regression) nor in the slope of the learning effect ($p = 0.680$, mixed-effects probit regression). A hierarchical Bayesian model using standard normal priors yielded evidence ratios favoring no difference in the intercept ($BF = 7.12$) nor in the slope of the learning effect ($BF = 3.94$). These results suggests it is unlikely that above-chance performance can be accounted for in terms of stimulus-based reinforcement learning.
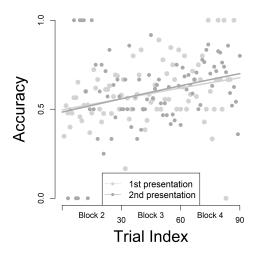

Figure 7: Learning over the course of the experiment was not attributable to stimulus-based reinforcement learning acting on sets of paired probe letters. The x-axis shows the trial number and the y-axis shows accuracy. Points show the group-level accuracy on that trial number broken into first and second presentations of paired probe letters. No significant differences in the intercept ($p = 0.815$, $BF = 7.12$, mixed-effects probit regression) or in the slope of the learning effect ($p = 0.680$, $BF = 3.94$, mixed-effects probit regression) were observed between the two sets of trials.

Discussion

We investigated how the timing of future events is cognitively represented. To do this, we developed the novel Judgment of Anticipated Co-Occurrence (JACO) paradigm, which required participants to predict the temporal overlap of future events. The task was designed to test whether future events are held in conjunction with their position on a prospective mental timeline, or whether they are represented with a strength-like signal indicating relative imminence. We further used response times to examine the mechanism by which the future is accessed, hypothesizing that participants would imagine the future as it unfolded through time and that response times would scale with the lag to the more imminent probe letter.

Participants' group-level performance in the task was significantly above chance, and a subset of participants performed significantly above chance as individuals. Successful task performance could not be explained by a strength-like estimate of future timing coupled with a linear decision process. Participants' behavior implied they were equipped with 1) temporally organized predictions of the future letters in the stream, 2) a temporal estimate of the lagged target window, and 3) some means to compare the two and evaluate their temporal alignment. Our results suggest that people can construct a timeline of near-future events based on their past experiences and subsequently access and reason about this constructed time-conjunctive future.

This result is inconsistent with the notion that predictions are made and acted upon using a scalar, strength-based estimate of future timing coupled with threshold-like decision process. Due to the placement of targets and lures, task performance using temporally discounted signal-strength would have been at chance level, as targets were equally likely to be more or less imminent. Reconciling our findings with a strength-like representation of temporal proximity would require a more complex, nonlinear decision process.

Even though full information to support task performance was available at the beginning of the testing phase, participant accuracy began near chance and increased across testing

trials. Before the testing phase began, participants had learned the sequence reasonably well and had been presented with the deterministic behavior of the lagged target window. One possible explanation for higher performance over time is that participants continued to build more accurate representations of the sequence and the lagged target window. Future studies might test this possibility by including longer periods of training or intermediate checks on the accuracy of participants' representations. Another possible explanation is that over time, participants gained facility with the evaluation of temporal co-occurrence. The improvement in accuracy over time did not appear to be attributable to stimulus-based reinforcement learning. Likewise, we regard it as implausible that participants succeeded at the task by independently learning pairwise associations between sequence positions and target items, given that each of the 45 position-target pairings occurred a maximum of three times in the experiment and such an account would not explain the observed effects of lure position on accuracy and response time.

Participant accuracy varied with the positioning of events in the future timeline, suggesting the existence of a temporal dimension along which events are located and differentiated. Accuracy was greater when the lure was further from the lagged target window (lures at lags 2 and 8) than when it was closer to the lagged target window (lures at lags 3 and 7). Accuracy was greatest for trials with lures at lag 2, suggesting that the nearest future was accessed with the most clarity. The placement of the target did not meaningfully affect accuracy, nor did whether the lure preceded or followed the target.

If a participant's imagined episodic future were sequentially accessed during trials, one would expect response times to increase in tandem with the temporal remoteness of the information necessary to support accurate decisions. That is precisely what was seen in the response time data, as response time increased with the lag to the more imminent probe letter. These results parallel findings in the memory and prospection literature, and are consistent with the notion of a self-terminating forward scanning process (Muter, 1979; Hacker, 1980; Hockley, 1984; McElree & Dosher, 1993; Singh & Howard, 2017).

Our response time data were not consistent with a theoretical "jump forward in time" similar to the "jump back in time" that characterizes episodic memory (Howard & Kahana, 2002). Response times were inconsistent with participants jumping forward in time to the lagged target window, the target region of interest. If such a process were used, we would not expect response times to increase with the lag to the more imminent probe letter; rather, we would expect them to remain constant or to vary with the lag to the target.

There are numerous questions still remaining about the flexibility of episodic prospection; variants of the present task could be used to test those hypotheses in future research. One outstanding question is whether the pattern of accessing the future depends on which temporal region of the future is relevant. One could run this task with a variety of different lagged target windows to test participants' ability to switch flexibly among them, and to test whether the dynamics of the forward scanning process remain the same at longer timescales. Another open question is whether the future is simulated continuously through time or as an ordinal succession of discrete events. To test this, one could deconfound time and ordinal position, which are interchangeable in the present version of the paradigm. Using temporally jittered item presentations or a continuously evolving stimulus would make it possible to determine whether the future is imagined continuously through time or as an ordinal succession of discrete events.

Our results demonstrate that people can craft a time-conjunctive representation of the near future informed by past experience, which has well-ordered temporal characteristics and is accessed sequentially. These results add to the existing body of evidence for a system of episodic prospection that mirrors episodic memory.

References

Abelson, R. P. (1981). Psychological status of the script concept. *American Psychologist*, **36** (7), 715–729.

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.

Atance, C. M., & O'Neill, D. K. (2001). Episodic future thinking. *Trends in Cognitive Sciences*, **5** (12), 533–539.

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

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, **67** (1), 1-48.

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

Bürkner, P.-C. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, **80** (1), 1-28.

Bürkner, P.-C. (2018). Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal*, **10** (1), 395-411.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software* **76** (1).

Champely, S. (2018). pwr: Basic Functions for Power Analysis. R package version 1.2-2. https://CRAN.R-project.org/package=pwr.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, **36** (3), 181–204.

Daniel, T. O., Stanton, C. M., & Epstein, L. H. (2013). The Future Is Now: Reducing Impulsivity and Energy Intake Using Episodic Future Thinking. *Psychological Science*, **24** (11), 2339-2342.

D'Argembeau, A., & Van der Linden, M. (2004). Phenomenal characteristics associated with projecting oneself back into the past and forward into the future: Influence of valence and temporal distance. *Consciousness and Cognition*, **13** (4), 844–858.

DeCarlo, L. T. (2012). On a signal detection approach to m-alternative forced choice with bias, with maximum likelihood and Bayesian approaches to estimation. *Journal of Mathematical Psychology*, **56** (3), 196-207.

Donkin, C., & Nosofsky, R. M. (2012). A power-law model of psychological memory strength in short- and long-term recognition. *Psychological Science*, **23** (6), 625-634.

Eklund, Aron (2016). beeswarm: The Bee Swarm Plot, an Alternative to Stripchart. R package version 0.2.3. https://CRAN.R-project.org/package=beeswarm.

Frederick, S., Loewenstein, G., & O'Donoghue, T. (2002). Time Discounting and Time Preference: A Critical Review. *Journal of Economic Literature*, **40** (2), 351-401.

Glenberg, A. M., Bradley, M. M., Stevenson, J. A., Kraus, T. A., Tkachuk, M. J., Gretz, A. L., Fish, J. H., & Turpin, B. M. (1980). A two-process account of long-term serial position effects. *Journal of Experimental Psychology: Human Learning and Memory*, **6** (4), 355-369.

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., & Maguire, E. A. (2007). Deconstructing episodic memory with construction. *Trends in Cognitive Sciences*, **11** (7), 299–306.

Hinrichs, J. V. (1970). A two-process memory-strength theory for judgment of recency. *Psychological Review*, **77** (3), 223-233.

Hintzman, D. L (2010). How does repetition affect memory? Evidence from judgments of recency. *Memory & Cognition*, **38** (1), 102-115.

Hintzman, D. L., & Block, R. A. (1971). Repetition and memory: Evidence for a multiple-trace hypothesis. *Journal of Experimental Psychology*, **88** (3), 297-306.

Hintzman, D. L., Block, R. A., & Summers, J. J. (1973). Contextual associations and memory for serial position. *Journal of Experimental Psychology*, **97** (2), 220-229.

Hockley, W. E. (1982). Retrieval processes in continuous recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **8** (6), 497-512.

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., & Kahana, M. J (2002). A Distributed Representation of Temporal Context.

Journal of Mathematical Psychology, **46** (3), 269-299.

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

Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory & Cognition*, **24** (1), 103–109.

Kierkegaard, S. (1938). *The Journals of Kierkegaard* (A. Dru, translator). London, UK: Oxford University Press (Original work written 1843).

Klein, S. B., Loftus, J., & Kihlstrom, J. F. (2002). Memory and temporal experience: The effects of episodic memory loss on an amnesic patient's ability to remember the past and imagine the future. *Social Cognition*, **20** (5), 353–379.

Kwan, D., Carson, N., Addis, D. R., & Rosenbaum, R. S. (2010). Deficits in past remembering extend to future imagining in a case of developmental amnesia. *Neuropsychologia*, **48** (11), 3179–3186.

Lemon, J. (2006). Plotrix: a package in the red light district of R. *R-News*, **6** (4), 8-12.

Liu, L., Feng, T., Chen, J., & Li, H. (2013). The value of emotion: how does episodic prospection modulate delay discounting? *PLoS One*, **8** (11).

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

Merkle, E., & You, D. (2018). nonnest2: Tests of Non-Nested Models. R package version 0.5-2. https://CRAN.R-project.org/package=nonnest2

Monsell, S. (1978). Recency, immediate recognition memory, and reaction time. *Cognitive Psychology*, **10** (4), 465-501.

Murdock, B. B., Jr. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, **64** (5), 482-488.

Murdock, B. B., & Metcalfe, J. (1978). Controlled rehearsal in single-trial free recall. *Journal of Verbal Learning & Verbal Behavior*, **17** (3), 309-324.

Murdock, B. B., & Okada, R. (1970). Interresponse times in single-trial free recall. *Journal of Experimental Psychology*, **86** (2), 263-267.

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

Navarro, D. J. (2015). Learning statistics with R: A tutorial for psychology students and other beginners. (Version 0.5). University of Adelaide. Adelaide, Australia.

Okuda, J., Fujii, T., Ohtake, H., Tsukiura, T., Tanji, K., Suzuki, K., Kawashima, R., Fukuda, H., Itoh, M., & Yamadori, A. (2003). Thinking of the future and past: The roles of the frontal pole and the medial temporal lobes. *NeuroImage*, **19** (4), 1369–1380.

Peters, J., & Büchel, C (2010). Episodic future thinking reduces reward delay discounting through an enhancement of prefrontal-mediotemporal interactions. *Neuron*, **66** (1), 138-148.

Peirce, J. W. (2007). PsychoPy - Psychophysics software in Python. *Journal of Neuroscience Methods*, **162** (1-2), 8-13.

Samuelson, P. (1937). A Note on Measurement of Utility. *Review of Economic Studies*, **4** (2), 155-161.

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** (5306), 1593-1599.

Shankar, K. H., Singh, I., & Howard, M. W. (2016). Neural mechanism to simulate a scale-invariant future. *Neural Computation*, **28** (12), 2594–2627.

Shepard, R. N., & Teghtsoonian, M. (1961). Retention of information under conditions approaching a steady state. *Journal of Experimental Psychology*, **62** (3), 302-309

Singh, I., & Howard, M. W. (2017). Scanning along a compressed timeline of the future. *bioRxiv*.

Spreng, R. N., Mar, R. A., & Kim, A. S (2009). The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode: a quantitative meta-analysis. *Journal of Cognitive Neuroscience*, **21** (3), 489-510.

Stan Development Team (2018). RStan: the R interface to Stan. R package version 2.18.2. http://mc-stan.org/.

Sutton, R. S., & Barto, A. G. (1981). Toward a modern theory of adaptive networks: Expectation and prediction. *Psychological Review*, **88** (2), 135-170.

Szpunar, K. K., Watson, J. M., & McDermott, K. B (2007). Neural substrates of envisioning the future. *Proceedings of the National Academy of Sciences of the United States of America*, **104** (2), 642-647.

Szpunar, K. K., & McDermott, K. B. (2008). Episodic future thought and its relation to remembering: evidence from ratings of subjective experience. *Consciousness and Cognition*, **17** (1), 330–334.

Tiganj, Z., Gershman, S. J., Sederberg, P. B., & Howard, M. W. (2019). Estimating scale-invariant future in continuous time. *Neural Computation*, **31** (4), 681-709.

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

Tulving, E. (1985b). Memory and consciousness. *Canadian Psychology*, **26** (1), 1–12.

Vuong, Q. H. (1989). Likelihood Ratio Tests for Model Selection and non-nested Hypotheses. *Econometrica*, **57** (2), 307–333.

Wang, T., & Merkle, E. C. (2018). merDeriv: Derivative Computations for Linear Mixed Effects Models with Application to Robust Standard Errors. *Journal of Statistical Software, Code Snippets*, **87** (1), 1-16.

Wickelgren, W. A. (1970). Multitrace Strength Theory. In D.A. Norman (Ed.), *Models of Human Memory*. New York: Academic Press, 65-102.

Williams, J. M. G., Ellis, N. C., Tyers, C., Healy, H., Rose, G., & Macleod, A. K. (1996). The specificity of autobiographical memory and imageability of the future. *Memory & Cognition*, **24** (1), 116–125.

Willingham, D. B., Nissen, M. J., & Bullemer, P. (1989). On the development of procedural knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **15** (6), 1047–1060.