Response dynamics of event-based prospective memory retrieval in mouse tracking

Jason L. Hicks¹ · Samantha N. Spitler¹ · Megan H. Papesh¹

Published online: 19 February 2019
© The Psychonomic Society, Inc. 2019

Abstract

Prospective memory (PM) is typically measured using keypresses in laboratory paradigms, which therefore assess only discrete, stage-like processes. In the present study we manipulated focal and nonfocal PM cue conditions, as well as participants' focus on different aspects of the PM/ongoing task set, using the methodology to capture dynamic computer mouse movements. The software captured mouse trajectories during lexical decisions and PM responses. We replicated many findings typical in the PM literature, including the accuracy advantage for focal over nonfocal conditions and longer ongoing-task response times for nonfocal conditions. Participants' movement trajectories during PM responses revealed evidence for both spontaneous-retrieval and strategic-monitoring processes in focal and nonfocal PM retrieval conditions. During trials suggestive of spontaneous retrieval, mouse trajectories initially went toward the typical ongoing-task response but turned mid-trajectory toward the PM response field on the opposite side of the computer screen. In nonfocal conditions, these trajectory reversals had a wider arc and took longer to complete, reflecting the likely greater retrospective retrieval requirements of nonfocal conditions. Regarding what are more likely to be strategic-monitoring processes, a significant portion of responses traveled directly to the PM response field, as though people were prepared to make such a response.

Keywords Prospective memory · Retrieval dynamics · Mouse tracking · Attentional monitoring · Spontaneous retrieval

Prospective memory (PM; Einstein & McDaniel, 1990) involves remembering to fulfill an intention in the future, such as remembering to deliver a message to a friend upon seeing her or remembering to stop at the grocery store on the way home from work. During such event-based PM, some stimulus in the environment cues an intention that needs to be fulfilled (e.g., seeing the grocery store sign). Event-based PM is typically studied in the laboratory by engaging participants in an ongoing task (e.g., lexical decisions) but giving them cues (e.g., certain words or classes of words) that, when encountered, signal the participant to suspend the primary task and fulfill a different intention (Einstein & McDaniel, 1990). For example, when participants encounter a PM cue (e.g., *horse*), they must recognize it as a cue and switch from making

word/nonword judgments to pressing a different key on the keyboard. The likelihood of remembering to fulfill an intention varies depending on the task, how well embedded the intention is, individual differences, and attentional processes (Einstein & McDaniel, 2005; Smith, 2003; Uttl, White, Gonzalez, McDouall, & Leonard, 2013).

Marsh, Hicks, Cook, Hansen, and Pallos (2003) specified four distinct stages in the processing of event-based PM cues encountered during some ongoing task: (a) recognizing the cue as related to an intention, (b) verifying that the cue and other aspects of the context meet the requirements for responding, (c) retrieving the appropriate response action, and (d) coordinating and executing the proper response to the PM cue, to the ongoing task, or both (see also Graf, 2005, for a similar perspective). As was discussed by Marsh et al. (2003), a major drawback of the typical computerized paradigm for measuring performance is that a single response to the PM cue represents the culmination of these processes, rather than allowing for distinctions among the processes or for one or more of the processes to be isolated. We propose that investigating mouse movements over time will afford a better ability to reveal one or more of these processing stages.

Electronic supplementary material The online version of this article (<https://doi.org/10.3758/s13421-019-00909-5>) contains supplementary material, which is available to authorized users.

✉ Jason L. Hicks
jhicks@lsu.edu

¹ Louisiana State University, Baton Rouge, LA, USA

Prospective memory and response dynamics

A relatively novel approach in PM is the use of mouse tracking, a methodology that records the x, y spatial coordinates from computer mouse movements over time (e.g., Hehman, Stoller, & Freeman, 2014; Scherbaum & Kieslich, 2018; Spivey, 2007; Spivey & Dale, 2006). One example of software that represents this methodology, which we used in our research, is MouseTracker (Freeman & Ambady, 2010). This software package samples information during mouse movements at rates from 60 to 75 times per second. Another example is Mousetrap (Kieslich & Henninger, 2017). Similar to eyetracking, mouse-tracking allows researchers to document dynamically unfolding cognitive events and the allocation of overt attention (Johnson, Mulder, Sijbinga, & Hulsebos, 2012). Unlike eye movements, which must be averaged over many trials to produce time-course estimates, mouse movements offer continuous data (Magnuson, 2005), and are often used when time-course information offers theoretical insight. For example, mouse trajectories have revealed the time course over which lexical competition is resolved, lending support to continuous, over discrete, models of language processing (Spivey, Grosjean, & Knoblich, 2005).

Mouse tracking has also been used to tap into real-time processing during social categorization tasks. For example, Freeman and Ambady (2010) presented participants with typical and atypical male and female faces on a computer screen for a male–female categorization task. After starting each trial with the mouse on a central screen location, participants moved the mouse and clicked the appropriate response field (MALE or FEMALE) for each face stimulus. Freeman and Ambady mapped the trajectories of participants' cognitive process on each trial and related these to psychological processes used when categorizing the face stimuli. They found that participants were typically “drawn” toward the opposite gender category when shown an atypical face, before ultimately deciding on the appropriate decision. For example, when shown a feminine male face, participants' physical movements were initially toward the FEMALE response, even though they ultimately clicked on the appropriate MALE response.

Mouse-tracking methodology was chosen because it affords the recording and analysis of many important variables. In MouseTracker specifically, variables we were interested in include area under the curve (AUC), maximum deviation (MD), time to maximum deviation (time to MD), and initiation time, among others (Freeman & Ambady, 2010). AUC is the geometric area between the trajectory made by the participant on a trial and an idealized straight-line trajectory toward the correct response field (Freeman & Ambady, 2010). Thus, trajectories that initially go toward the wrong response field (e.g., on the left side of the screen), but ultimately reverse and end on the correct response field (e.g., on the right side of the

screen) would result in a high AUC value. MD is similar to AUC but measures the maximum perpendicular deviation from the idealized straight-line trajectory toward the correct response field. As a result, AUC and MD are highly correlated. Time to MD is the time in milliseconds it takes participants to reach MD. Finally, initiation time is the time in milliseconds that it takes for participants to move the mouse off the Start response field at the beginning of a trial.

We propose that measures of dynamic responding will afford new quantitative and qualitative ways of characterizing PM retrieval. For example, at what point following the presentation of a PM cue does one realize that it is relevant to an intention and requires a deviation from the ongoing-task response (e.g., the noticing or verification stage)? Do people generally slow down their initiation of responses to all stimuli in the ongoing task, as might be the case if people are more cautious not to miss PM cues (Heathcote, Loft, & Remington, 2015)? To answer such questions, we implemented a standard focal versus nonfocal manipulation of PM cues to measure how mouse tracking measures correlate with different PM intentions. Focal tasks are those in which processing required for the ongoing task overlaps substantially with processing required to notice a stimulus as related to an intention (Einstein & McDaniel, 2005). Nonfocal tasks are those in which little such overlap exists, and they generally require more attentional resources to realize that a stimulus is intention related. It is well established that responses to non-PM stimuli are generally slower when participants have been given nonfocal PM instructions (e.g., respond to any animal words; Hicks, Franks, & Spitler, 2017; Marsh et al., 2003) as opposed to focal PM instructions (e.g., respond to the specific word *horse*). Mouse tracking may allow us to isolate where, in the flow of information processing, this slowing occurs. Do people hesitate before initiating their movement, or do they make slow, more cautious ongoing movements? Detection of focal PM cues is also generally better than detection of nonfocal PM cues. Examining mouse trajectories during PM trials will help us to more precisely isolate the time course of detection. For example, focal PM retrieval is often interpreted as the result of spontaneous noticing processes (Einstein & McDaniel, 2005), which could relate to measurable differences at earlier stages of responding, relative to nonfocal trials. During PM trials, this may manifest as an immediate and direct path to the PM response option. Alternatively, because spontaneous noticing is theorized to occur without prior attentional monitoring, such noticing may occur mid-trajectory, while people are making a typical ongoing-task response. Yet such noticing should still occur more rapidly than retrieval in a nonfocal context, because of the additional categorical retrieval demand of our nonfocal condition.

In addition to manipulating focal versus nonfocal PM cues, we used an emphasis instruction as another way of influencing

PM retrieval. Some participants were told the importance of responding to the PM cues correctly, others were told to emphasize responding quickly and accurately to the ongoing task, and still others were not given emphasis instructions. Emphasizing the importance of the PM task improves PM accuracy in time-based and event-based PM tasks (Kliegel, Martin, McDaniel, & Einstein, 2001, 2004). Moreover, higher importance on the PM task should increase the likelihood of a conscious monitoring strategy and increase response times (RTs) generally (e.g., Horn & Bayen, 2015), but at which stage in the ongoing decision remains unknown (e.g., before the mouse movement begins vs. after).

To our knowledge, one prior published study has used mouse tracking to investigate the dynamics of PM attentional monitoring (Abney, McBride, Conte, & Vinson, 2015; see also Kurtz, 2017). Abney et al. measured the response dynamics of ongoing lexical decision trials while participants anticipated PM cues. Responses to detected cues were issued by a keypress following the mouse-driven ongoing-task response (i.e., participants did not abandon the primary task, but pressed the “z” key after clicking their lexical decision response with the mouse). Abney et al. measured changes in movement velocity with mouse tracking and found that successful detection of PM cues was associated with a delayed maximum velocity on correct ongoing-task RTs, relative to missed PM cue trials. They interpreted this pattern as evidence of PM cue interference, such that successfully detecting the PM cue causes interference with the required ongoing-task response (Marsh et al., 2003; Marsh, Hicks, & Watson, 2002). In addition, peak velocities of non-PM cue lexical decision trials for participants given a nonfocal PM instruction were higher than those given a focal PM instruction, which they interpreted as evidence of more controlled monitoring associated with nonfocal PM conditions. Initiation times did not differ for focal versus nonfocal conditions, although the numeric trend was for initiation times to be faster for focal trials that generated successful PM detection. Finally, they found a nonsignificant difference in accuracy to detect PM cues between the focal and nonfocal PM conditions, although the trend favored the usual advantage for focal conditions (e.g., Marsh et al., 2003).

A significant change in our methodology is that PM responses were required to replace the regular lexical decision response, and we also required a mouse-driven PM response in a different region of the two-dimensional layout, rather than requiring the PM response with a keypress following the LDT mouse response (cf. Abney et al., 2015). We used a standard lexical decision task (LDT) for our ongoing task, in which participants made word/nonword judgments by clicking the appropriate response field (see Fig. 1). The WORD and NONWORD response fields were each to the right of the screen-centered lexical decision stimulus. Sham fields were visible on the left side of the screen, with the

upper-left field being relevant to the PM response instructions given to participants. Placing the PM response field on the opposite side of the screen from the WORD/NONWORD fields, but vertically in the same position as the WORD response field, created a situation in which the typical response had to change drastically in order to promote accurate PM retrieval, but would thus provide evidence of curvature in the trajectory. The positioning of the WORD and PM response fields also distinguished responses that traveled more directly to the PM response field from those that began toward the WORD field and reversed back toward the PM field. Thus, this feature had the added benefit of showing obvious qualitative and quantitative changes in the response dynamics on PM cue trials. As we suggested earlier, one possibility is that because focal PM cues are theorized to occur via spontaneous-retrieval processes without prior attentional monitoring, the typical LDT response would be interrupted by a reversal in trajectory from right to left, as the spontaneous retrieval occurred in the middle of stimulus processing. In addition, because focal spontaneous retrieval may occur more quickly in stimulus processing than in nonfocal retrieval, the measures of MD, time to MD, and AUC should be smaller for focal than for nonfocal PM cue trials. A complementary prediction was that response trajectories might initially travel more directly to the PM response field if people were more consistently monitoring for PM cues. This type of direct responding might be more prevalent following nonfocal instructions and for PM task emphasis instructions. One might also expect movement initiation times to be slower for trials associated with more attentional monitoring, because people might wait to verify a stimulus as being intention related prior to movement.

During the ongoing lexical decision trials (i.e., those in which PM cues were not presented), overall RTs to words should be longer for people given a nonfocal instruction (e.g., Hicks et al., 2017; Marsh et al., 2003). Whether this slowing affects early or late stages of processing is an open question. Abney et al. (2015) found that the maximum velocity for regular lexical decision trials occurred later in time for people given nonfocal instructions, relative to those with focal instructions. This would be represented by a longer time to MD in response trajectory curvature. In contrast, if people in nonfocal conditions are relatively more cautious in responding on all trials (Heathcote et al., 2015), then the initiation of a mouse movement upon seeing the stimulus might be slower, with the remaining RT for mouse movements being no different for focal versus nonfocal conditions. Finally, the initiation time might depend on whether the importance of the PM task has been emphasized, as opposed to an LDT emphasis. Presumably, more important intentions will prompt caution in very early stages of the decision, and therefore might slow down initiation times.

Fig. 1 Depiction of the computer screen layout for each trial of the lexical decision task. Participants had to place their mouse cursor over the START field (left screen in the figure). Clicking the left mouse button initiated the appearance of the stimulus screen with its associated

response fields (right screen in the figure). The Xs represent tape on the tabletop for participants to settle their mouse onto at the beginning of each trial, depending on preferred handedness

Method

Participants

Undergraduate psychology students from Louisiana State University participated in the study for partial class credit or extra credit. All participants were native English speakers. Participants were tested individually in sessions lasting approximately 30 min. A total of 265 participants were randomly assigned to seven conditions. Six of the conditions, ranging in sample size from 37 to 41, formed a factorial crossing of task emphasis (PM, LDT, or standard) with focality (focal or nonfocal). We tested an additional no-intention control condition ($n = 37$).

Materials

A total of 130 medium-frequency words (four to ten letters long, HAL frequency 3,000 to 14,000, $M = 10,562$) and 130 pronounceable nonwords were chosen from the English Lexicon Project (Balota et al., 2007). Each participant completed three blocks, with words and nonwords randomly presented. The first block consisted of 20 practice trials (ten words, ten nonwords). The second and third blocks consisted of 120 trials each (60 words, 60 nonwords). Only the third block had prospective cues, with the focal cue (the word *animal*) or nonfocal cues (each of the words *horse*, *monkey*, *rabbit*, *tiger*, *elephant*, and *chicken*) appearing on Trials 18, 36, 54, 72, 90, and 108 of that block.

Participants' task emphasis was manipulated in the focal and nonfocal conditions by highlighting different aspects of their performance. To emphasize PM performance, the instructions stressed the importance of detecting the PM targets (e.g., for the focal condition, *It is very important that you*

press the TOP LEFT button whenever you see the word "ANIMAL"; for the nonfocal condition, It is very important that you press the TOP LEFT button whenever you see any "ANIMAL WORD"). To emphasize ongoing-task performance, the instructions stressed the importance of the LDT (e.g., *It is very important that you make accurate word/nonword decisions*). Additionally, one group each within the focal and nonfocal conditions did not receive instructions emphasizing one aspect of performance (i.e., they received the standard instructions). Finally, a seventh condition served as a no-intention control, in which each block consisted of a LDT with no importance manipulation or mention of an intention.

Procedure

Participants first provided informed consent. The instructions for the initial practice block were presented on the computer screen, and participants read the instructions aloud (throughout the experiment, participants read all instruction screens aloud). To begin each trial, participants positioned the computer mouse over an "X" marked on the workstation and clicked the START response field on the computer screen. As is shown in Fig. 1, a word or nonword then appeared above the START field, printed in uppercase 12-point font, and four response fields (symmetrically at each corner of the computer screen) became visible. The WORD field was in the top right corner, the NONWORD field in the bottom right corner, and two sham fields were in the top/bottom left corners. Participants made lexical decisions by clicking the WORD or NONWORD field, and identified PM targets by clicking the top-left sham field. Throughout each response, the x -, y -coordinates for the mouse cursor were sampled at 70 Hz (Freeman & Ambady, 2010). To encourage participants to

initiate their responses quickly, trials with movement initiation times longer than 500 ms were followed by the message “You moved the mouse too late.” There was a 1,000-ms interval between each trial.

A second block of trials consisted of word/nonword decisions as described above, with 120 trials total. Prior to beginning the third block, participants’ experience differed on the basis of their condition assignment. For the focal conditions, participants were instructed to click in the upper-left response field when the word ANIMAL appeared. Participants in the focal condition with a PM emphasis were further instructed that it was very important that they remember to click the top-left response field when the word ANIMAL was presented. Those in the focal condition with the LDT emphasis were told that it was very important that they make accurate word/nonword decisions. The third focal condition did not receive an emphasis instruction (standard).

For the three nonfocal conditions, participants were instructed to click the response field in the top-left corner when any animal words appeared—for example, “DOG” (which was not one of the cues shown). The same emphasis instructions (standard, LDT, and PM emphasis) were given to separate nonfocal groups. The control group (seventh condition) was instructed to make word/nonword decisions as they had done in the previous blocks. They were not given an intention otherwise.

Upon completion of the three blocks, participants were thanked for participating and asked to complete a postquestionnaire. On the postquestionnaire, participants were asked to reflect on how well they think they did, to recall what the overall task was, to recall the PM cues, and to recall the appropriate response to the PM cues.

Results

Data analysis was conducted using IBM SPSS version 23. Prior to inferential analysis, exclusion criteria were applied in sequence. First, all trials with mouse initiation times greater than 500 ms were excluded. This affected 3.5% ($n = 40$) of all PM cue trials and 8.1% ($n = 2,375$) of LDT word stimulus trials. Second, PM cue trials were considered outliers if their trajectories either entered all four quadrants of the x, y space on the screen or entered the response-irrelevant bottom-left quadrant of the space. These responses ($n = 98$) represented 8.3% of all PM cue trials. Third, RT exclusion for correct LDT word trials was conducted for each participant, such that trials longer than 2.5 standard deviations (SDs) from the participant’s mean were excluded. This resulted in a further 2.3% of word trials being excluded across Blocks 2 and 3 ($n = 620$). Fourth, participants with less than 60% correct on the ongoing LDT in either Block 2 or Block 3 were excluded ($n = 6$). Fifth, participants with a large proportion of correct but spatially

Table 1 Average proportions of correct PM responses across focality and task emphasis conditions

Emphasis	Focality						Average	
	Focal			Nonfocal				
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Standard	35	.89	.20	33	.67	.34	.78	.28
PM	35	.90	.15	32	.70	.25	.80	.21
LDT	36	.74	.36	38	.60	.38	.67	.37
Average		.84	.26		.66	.32		

outlying “word” responses in either Block 2 or Block 3 of the LDT were excluded ($n = 5$). These were defined as anyone with over 50% of correct trials that entered the portion of the response screen left of the START field (i.e., x -coordinates less than -0.15). Sixth, participants were excluded if their median RT was more than 2.5 SDs beyond their condition mean for either PM direct or reversal trajectories ($n = 4$). Seventh and finally, further participants were excluded if their Block 2 or Block 3 trials were more than 2.5 SDs beyond their condition mean for correct LDT word direct or reversal trajectories ($n = 4$). Thus, 19 participants in total were excluded on the basis of these collective criteria. We note here that our results were not affected by these exclusion criteria. The primary outcomes were consistent even when no trials and no participants were excluded (see Appx. 1).¹ The resulting sample sizes for each of the primary conditions following these exclusions are listed in Table 1. For all analyses, the Type I error rate was set to .05. Post hoc comparisons were conducted with a per-comparison Ryan adjustment to the Type I error rate.

PM cue trials: Accuracy

To analyze overall PM accuracy in the third LDT block, we conducted a 2×3 between-groups analysis of variance (ANOVA) with focality (focal vs. nonfocal) and emphasis (PM task emphasis, LDT emphasis, standard emphasis) as fixed factors. Descriptive statistics are presented in Table 1. The main effect of PM focality was significant, $F(1, 205) = 19.92$, $MSE = 0.09$, $p < .001$, $\eta_p^2 = .09$. Participants in the focal condition ($M = .84$) were more accurate than those in the nonfocal condition ($M = .66$). The main effect of PM emphasis was also significant, $F(2, 205) = 4.53$, $p < .05$,

¹ In the review process, this amount loss of data due to exclusions was brought up as a concern. We argue that our exclusions were done on a principled basis, to ferret out unusual responses (e.g., mouse trajectories that entered multiple quadrants of space) or people that were not doing the task well, and to bring the sample distributions of the data closer to the expectations of normality and variance homogeneity. Importantly, Appendix 1 reveals that our outcomes did not depend on the exclusions we applied.

Fig. 2 Averaged time-normalized response trajectories for correct PM responses by trajectory type (direct vs. reversal) and focality (focal vs. nonfocal). The START field is represented at the bottom of the figure but

was shown in the middle of the computer screen, as displayed in Fig. 1. The PM response field is in the upper left. Not shown is the WORD response field in the upper right

$\eta_p^2 = .04$. Post hoc analyses showed that the PM emphasis condition ($M = .81$) reliably differed from the LDT emphasis condition ($M = .67$), but not the standard emphasis condition ($M = .78$). The standard and LDT emphasis conditions did not differ statistically. The interaction was not significant.

PM cue trials: Response dynamics

Prior to analysis, the computer display was divided into four quadrants based on the x -, y -coordinate location of the START field. The boundary of the START field was located at x -coordinates $-.15$ (left) to $.15$ (right) and y -coordinates $.65$ (bottom) to $.85$ (top). For PM trials, trajectories were divided into four types: direct, reversal, outlier, and PM failures. Accurate PM target trajectories that entered the upper-left quadrant from the START field (x -coordinates $< -.15$ and y -coordinates $> .85$) were labeled *direct PM trajectories* (25.2% of PM trials). Accurate PM target trajectories that entered the upper-right quadrant (x -coordinates $> .15$ and y -coordinates $> .85$) and then reversed back to the upper left were labeled *reversal PM trajectories* (49.8% of PM trials). Accurate trajectories on PM trials that entered into both quadrants on the left side of the screen (upper and lower left) or that entered all four quadrants were excluded as outliers, as we described earlier (8.3% of PM trials). The remaining PM cue trials were inaccurate responses (i.e., PM failures). See Fig. 2 for a visualization of the averaged direct and reversal trajectories on PM target trials across the focality factor.

The proportions of PM cue trials resulting in hits were analyzed with a $2 \times 2 \times 3$ mixed-factor ANOVA with trajectory type (direct vs. reversal), focality, and emphasis as

factors.² The main effect of trajectory type was significant, $F(1, 205) = 64.98$, $MSE = .10$, $p < .001$, $\eta_p^2 = .24$, in that the proportion of PM trials resulting in direct hits ($M = .33$) was less than the proportion that were reversals ($M = .42$). The main effect of focality was confirmed, $F(1, 205) = 19.92$, $MSE = .043$, $p < .01$, $\eta_p^2 = .09$, with better PM performance in the focal condition. The main effect of emphasis was also significant, $F(2, 205) = 4.53$, $p < .05$, $\eta_p^2 = .04$, replicating the overall analysis of PM accuracy that did not differentiate trajectory types in the ANOVA model. The PM emphasis and LDT emphasis conditions were significantly different in a post hoc analysis, as were the standard and LDT emphasis conditions. Only the standard-versus-PM emphasis comparison was not significant. No interactions were significant, all F s < 1.0 . Thus, the overall effects of focality and emphasis replicated when trajectory type was added as a factor.

Recall that initiation times reflect the time it takes for participants to begin moving the mouse following the onset of a stimulus. The only significant effect from the $2 \times 2 \times 3$ mixed-factor ANOVA on successful median PM trial initiation times was that of trajectory, $F(1, 136) = 15.37$, $MSE = 10,414.45$, $p < .001$, $\eta_p^2 = .10$. People took longer to initiate mouse

² The degrees of freedom in analyses involving trajectory type as a within-subjects factor are reduced because some participants were missing either direct trajectories or reversal trajectories, or both, among their PM responses, and therefore did not enter into the analyses. For the focal conditions, 20 participants produced reversals only, five participants produced direct trajectories only, and a further seven produced neither, representing a loss of 32 participants. The numbers in nonfocal conditions were roughly similar in proportion, with 25, three, and 13 participants, respectively, representing a total loss of 41. These were distributed in roughly equal proportions across the emphasis conditions, although the nonfocal/LDT emphasis condition lost the most, at 20. Thus, the resulting sample sizes across the 3×2 conditions ranged from 19 to 26.

movements on trials associated with direct PM trajectories ($M = 318$) than on trials associated with reversal PM trajectories ($M = 271$). No other effects were significant, all $ps > .15$.

The median times to MD of the trajectories from idealized straight lines also showed influences of both trajectory type and focality. In the $2 \times 2 \times 3$ mixed-factor ANOVA, these two factors interacted, $F(1, 136) = 9.62$, $MSE = 5,552.39$, $p < .01$, $\eta_p^2 = .07$. Simple main-effect tests demonstrated that the time to MD for reversals was longer for nonfocal than for focal conditions, $F(1, 136) = 56.99$, $MSE = 5,659.69$, $p < .001$, $\eta_p^2 = .31$. Nonfocal direct PM responses also took longer to reach MD than focal direct PM responses, but with a much smaller effect size, $F(1, 136) = 8.00$, $MSE = 8,202.94$, $p < .01$, $\eta_p^2 = .06$. The remaining measures associated with trajectory curvature on PM trials—AUC and MD—produced similar interactions of trajectory type with focality. In each case, significant differences were found between nonfocal and focal conditions only for the reversal PM trials, not for the direct PM trials.³ The main effect of emphasis condition was not significant for any of these measures. Summary statistical information is reported in Table 2.

Overall median RTs for correct PM responses were analyzed with a similar $2 \times 2 \times 3$ mixed-factor ANOVA. The main effect of trajectory type (direct vs. reversal) was significant and substantial in effect size, $F(1, 136) = 191.83$, $MSE = 13,928.42$, $p < .001$, $\eta_p^2 = .59$, but it was qualified by an interaction with focality, $F(1, 136) = 8.76$, $p < .01$, $\eta_p^2 = .06$. Simple main-effect tests demonstrated that reversal RTs were longer for nonfocal than for focal conditions, $F(1, 136) = 24.21$, $MSE = 17,266.97$, $p < .001$, $\eta_p^2 = .15$, but no such difference emerged for direct PM responses, $F(1, 136) = 1.18$, $MSE = 19,695.30$, $p = .28$, $\eta_p^2 = .009$. The more apparent difference in curvature seen in Fig. 2 between focal and nonfocal reversal trajectories, but not for direct trajectories, supports this analysis. The two-way interaction between focality and emphasis was marginal, $F(2, 136) = 2.86$, $MSE = 23,303.85$, $p = .061$, $\eta_p^2 = .04$. Nonfocal RTs were longer than focal RTs in the standard emphasis condition, $F(1, 136) = 16.30$, $MSE = 11,516.93$, $p < .001$, $\eta_p^2 = .11$, but not for the PM emphasis or LDT emphasis conditions, smallest $p > .19$ (largest $\eta_p^2 = .01$).

Non-PM cue LDT trial response dynamics

Recall that in Block 2 of the LDT, the intention had not yet been given to participants. Thus, comparing Block 2 to Block 3 performance tests would show whether the Block 3 intention interfered with LDT performance. The no-PM control group served as an additional control, since no intention was

Table 2 Simple main effects of focality for direct and reversal trajectory types, for various mouse-tracking measures and the overall response times on correct PM trials

	Focal		Nonfocal		Row simple effect	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>p</i>	Partial η^2
Response time (RT)						
Direct	991	130	1,016	103	.28	.01
Reversal	1,144	148	1,254	160	< .01	.15
Area under curve (AUC)						
Direct	0.06	0.07	0.07	0.07	.58	< .01
Reversal	0.83	0.39	1.09	0.36	< .01	.10
Maximum deviation (MD)						
Direct	0.10	0.10	0.11	0.10	.56	< .01
Reversal	0.64	0.21	0.81	0.18	< .01	.14
Time to MD						
Direct	534	86	577	99	< .01	.08
Reversal	544	61	642	101	< .01	.31

All measures are averaged medians (analyses of the participant means showed similar results). RT and time to MD are measured in milliseconds. The two-way interaction between focality and trajectory type was significant for each of these measures, and p values are not corrected for family-wise Type I error inflation in this table. Refer to Fig. 2 for the response trajectories associated with these data

provided in either block. We conducted a 2×7 mixed-factor ANOVA with block as a repeated measures factor and group as a between-subjects factor, to examine accuracy for correctly identifying word stimuli in the LDT. Only the main effect of block was significant, $F(1, 234) = 31.80$, $MSE = .001$, $p < .001$, $\eta_p^2 = .12$, all other $ps > .09$. Accuracy was better in Block 2 ($M = .986$) than in Block 3 ($M = .974$).

Regardless of group (including the no-PM control), in Block 2 direct trajectories for correct “word” responses were most frequent ($M = .81$), with fewer reversals ($M = .15$), and even fewer outliers ($M = .03$). Incorrect “nonword” responses were extremely low ($M = .02$). In Block 3, these respective means were .76, .17, .04, and .03. Thus, as LDT accuracy dropped slightly in Block 3, so did the relative proportion of direct to reversal trajectories. This latter observation was confirmed in a $2 \times 2 \times 7$ mixed-factor ANOVA with block and trajectory type (direct vs. reversal) as repeated measures factors and group as a between-subjects factor. The main effect of block, $F(1, 234) = 14.35$, $MSE = .002$, $p < .001$, $\eta_p^2 = .06$, was qualified by a Block $\times$ Trajectory interaction, $F(1, 234) = 29.51$, $MSE = .01$, $p < .001$, $\eta_p^2 = .11$. Relative to Block 2, Block 3 produced fewer correct direct trajectories ($p < .001$, $\eta_p^2 = .13$) but more reversals ($p < .001$, $\eta_p^2 = .07$). It is important to note that this pattern of a drop in direct trajectories and an increase in reversal trajectories across blocks occurred for all groups, including the no-PM control, which indicates a general influence of block rather than an effect specific to the maintenance of an intention in Block 3 for the PM groups.

³ When these simple main effect analyses were performed separately for direct and reversal trajectories, to allow for maximum statistical power (i.e., instead of the listwise deletion of trajectory data; cf. note 1), the outcomes and interpretation did not change.

Next, we analyzed RT and trajectory curvature information, including initiation time, time to MD, AUC, and overall RT. Block 3 median RTs for correct word responses for the PM groups were analyzed separately for direct ($M = 966$ ms across all intention conditions) and reversal ($M = 1,240$ ms) responses. Moreover, because Block 2 RTs for these measures did not differ statistically across conditions, but correlated with Block 3 RT ($r = .84$ for direct trajectories and $r = .54$ for reversal trajectories), we used Block 2 RT as a covariate in our analyses. For direct trajectories, a 2×3 analysis of covariance (ANCOVA) with focality and emphasis as fixed factors produced a main effect of focality, $F(1, 199) = 11.73$, $MSE = 3,259.18$, $p < .001$, $\eta_p^2 = .06$, with nonfocal conditions ($M = 978$ ms) being slower than focal conditions ($M = 951$ ms). A similar pattern was shown for median RTs on reversal trials, with a main effect of focality, $F(1, 179) = 4.69$, $MSE = 23,220.64$, $p < .05$, $\eta_p^2 = .026$, in which the nonfocal reversal RT ($M = 1,248$) was slower than the focal reversal RT ($M = 1,199$).⁴ Moreover, a similar ANCOVA revealed that the time it took direct trajectories to reach MD was slower for nonfocal participants ($M = 536$ ms) than for focal participants ($M = 516$ ms), $F(1, 199) = 12.02$, $MSE = 1,722.35$, $p < .001$, $\eta_p^2 = .06$. The same pattern was true for time to MD on reversal trials, $F(1, 179) = 6.22$, $MSE = 9,913.90$, $p < .05$, $\eta_p^2 = .034$, which was longer for nonfocal participants ($M = 640$ ms) than for focal participants ($M = 603$ ms). No significant effects were found for initiation times, either for direct trajectories, smallest $p > .74$, or for reversal trajectories, smallest $p > .19$.

Consistent with the RT analyses, measures of AUC and MD for Block 3 word trials were clearly larger for reversal than for direct responses, producing η_p^2 measures of effect size greater than .80 with trajectory type as a factor, but neither measure differed by focality or by emphasis for either direct or reversal trajectories, all $ps > .30$. Descriptive statistics for these measures related to accurate non-PM cue “word” responses in the LDT are available in Appendix Table 5.

Discussion

There are four primary outcomes to summarize with regard to PM retrieval response dynamics. First, the majority of correct PM responses were characterized as reversals, in which the

mouse initially traveled toward the WORD response field on the computer screen but then reversed along the x -axis, to land on the PM response field. This suggests that people sometimes retrieved the intention after initiating the manual “word” response otherwise required by the LDT. A sizeable proportion of trials, however, were direct responses, in which people traveled directly from the Start field to the PM response field. Second, the time to initiate direct PM responses was significantly longer than the time to initiate reversals, which may indicate increased attentional monitoring or preparatory attention related to the intention on direct trials. This difference in direct versus reversal initiations did not depend on focality or emphasis. Third, the proportion of reversals to direct trajectories was not moderated by focality or by task emphasis. One might have predicted that nonfocal trials would produce relatively more direct trajectories, under the assumption that people are more likely to actively monitor for intention-related stimuli in the LDT (i.e., to be ready to make a PM response). Fourth, the reversal response characteristics were different for the nonfocal and the focal conditions. Nonfocal reversal trajectories had a larger AUC (see Fig. 2), had a longer RT, and were in x - y -coordinate dimensions farther away from the idealized straight path (i.e., MD) than were focal reversals. Direct trajectories were less affected by focality, with only the time to MD being longer for nonfocal than for focal conditions.

One reviewer of our work was appropriately concerned that the nature of our focality manipulation might have affected our results. The concern was that our focal conditions presented a single PM cue six times over the course of the ongoing task, whereas the nonfocal condition presented six different animal words as the PM cues. Thus, stimulus repetition over the course of the experiment might have altered the focal relative to the nonfocal condition. We appreciate this concern, and in fact have published data from our laboratory directly focused on this issue (Hicks et al., 2017). As such, we examined mouse trajectory information for the first PM trial only as a comparison. The essential patterns of the data that occurred for reversal trials all replicated: Nonfocal RT was longer, nonfocal AUC and MD were larger, and time to MD was longer for nonfocal reversal than for focal reversal trials. Although only the overall RT and time to MD metrics created significant statistical differences, the other metrics were in the same direction, but had more significant variability in the data (i.e., the mean differences were very similar, but the standard deviations were 1.5 times larger). The lack of a difference in focal versus nonfocal initiation times for reversals was also replicated. Moreover, even the slight but significant difference for focal versus nonfocal direct-trial times to MD was replicated. Thus, our overall results reported in Table 2 are consistent with behavior for the first PM trial only.

Our results concerning PM trial response dynamics highlight the ambiguity in the examination of RT from keypress PM paradigms. In the typical case, when researchers use an

⁴ We also showed a nominal slowing of nonfocal direct responses as compared to the no-PM control group direct responses, in the context of planned comparisons of the no-PM control ($M = 956$ ms) with the focal standard condition ($M = 961$ ms), and of the no-PM control with the nonfocal standard condition ($M = 983$ ms), using Block 2 RT as a covariate. The latter contrast was significant, $t(97) = 1.83$, one-tailed $p = .035$. This replicated the standard task interference found in prior work for nonfocal but not for focal conditions, using only the standard condition as a reference. A similar analysis of reversal trials showed no significant differences, although interestingly, the focal standard condition was nominally faster ($M = 1,228$ ms) than the no-PM ($M = 1,240$ ms) and nonfocal standard ($M = 1,276$ ms) conditions.

ongoing task in which the PM keypress response replaces the ongoing-task response (e.g., Einstein & McDaniel, 1990), this change requires using some third key, which confounds manual response movement with the ongoing-task and intention-retrieval processes. A compromise is to ask people to make the ongoing-task response and then the PM response in a subsequent phase of the trial, which allows one to measure ongoing-task slowing due to actual PM retrieval, but this renders direct RT analysis of the PM behavioral response extremely difficult (e.g., Abney et al., 2015; Marsh et al., 2003; Marsh et al., 2002). Mouse-tracking methodology allows for a real-time analysis of how PM retrieval occurs, and at what time point during the flow of cognitive processing.

We found novel effects in response dynamics for ongoing LDT trials as well. Replicating prior work (e.g., Marsh et al., 2003; Smith, 2003) people took longer to make correct “word” responses in nonfocal, relative to focal, conditions in Block 3 of the ongoing task when the intention was active. This was true for direct LDT “word” responses and nominally true, but nonsignificant, for reversal responses. Although mouse movement responses are slower than keypress RTs (which typically range between 600 and 700 ms), the same behavioral patterns were observed and more fine-grained response patterns could be analyzed. For example, we observed that the time to MD both for direct and reversal word trajectories was slower for nonfocal relative to focal conditions, which may reflect general increased attentional monitoring on non-PM cue trials. This is consistent with an analogous nonfocal slowing on the direct PM responses: Overall RT slowing in ongoing-task performance in nonfocal conditions was revealed as a slightly longer time to reach maximum deviation, rather than a difference in initiation time.

Response dynamics as related to theories of event-based PM retrieval

We suggest that the real-time mouse movements during PM trials and non-PM trials generally support the two mechanisms posited to underlie PM performance according to the multiprocess framework (Einstein & McDaniel, 2005; McDaniel & Einstein, 2000) and the more recent dynamic multiprocess framework (Scullin, McDaniel, & Shelton, 2013): Retrieval is sometimes supported by attentional monitoring and sometimes by spontaneous retrieval processes, even across trials for a given person and condition (Scullin et al., 2013). Attentional monitoring is typically implicated when the ongoing-task RT or other metrics of ongoing-task performance are impaired while an intention is active (Marsh et al., 2003; Smith, 2003). Nonfocal conditions have typically been used to model conditions in which attentional monitoring is required for successful PM performance. In contrast, spontaneous retrieval occurs when reflexive processes support successful PM performance, notably even when attentional

processes are not evident. Subjective experience of spontaneous retrieval is associated with feelings that the intention just “popped” into mind (e.g., Meier, Zimmermann, & Perrig, 2006).

In the case of mouse tracking dynamics, because direct trajectories for successful PM trials were initiated more slowly than reversal trials, they may reflect attentional preparation consistent with a resource-demanding process, potentially even a target-checking type of process on those trials (Guynn, 2003; Smith, 2003). Moreover, differences in focality for overall RT of direct PM responses were not significant, suggesting that these direct responses may be associated with preparatory attention regardless of focality. Conversely, we suggest that reversal trajectories reflect retrieval that occurs after a regular ongoing-task response has been initiated, perhaps spontaneous in nature for the focal condition. These reversals were different in character for nonfocal as opposed to focal conditions, in that the movements took longer to complete and had a wider arc (see Fig. 2). That difference likely reflects a difference in faster specific-word retrieval in focal conditions versus slower categorical retrieval in nonfocal conditions. Such retrieval should not be necessary if the conditions of the intention are already in mind during conscious attentional monitoring (i.e., direct responses).

These differences in mouse trajectories may generally reveal the utility of the mouse-tracking methodology in revealing the microstructure of PM retrieval processes over the course of a trial (Marsh et al., 2003; Rummel, Wesslein, & Meiser, 2017). As was discussed by Marsh et al. (2003), a given trial producing successful retrieval should reflect multiple processes, including noticing that a cue is perhaps related to an intention, verification of this relationship, and retrieval/coordination of the proper behavioral response. Mouse trajectories likely help to isolate one or more of these processes from others within an overall RT. One strong possibility is that time to MD reflects differences in verification of the stimulus as related to an intention, which should take longer for nonfocal than for focal retrieval, because our nonfocal condition was categorical in nature. It is possible also that time to MD may reflect general familiarity-based cue noticing processes. Manipulating response instructions or the set size of specific cues may help tease apart whether noticing, versus cue verification or retrospective retrieval of contents, affects various mouse-tracking measures.

Surprisingly, the relative proportion of direct to reversal PM trajectories was not affected by the focality and emphasis manipulations. One might have predicted that if focal conditions are more likely to promote spontaneous retrieval, with a concomitant drop in attentional resource demands, then they might have produced a greater proportion of reversal trajectories than would nonfocal conditions. In contrast, all conditions produced direct versus reversal trajectories at roughly a 1:2 ratio. This suggests that even for people in nonfocal

conditions, a major contribution to their PM retrieval came from individual trials in which the intention popped into mind after the ongoing-task response was initiated. This could be consistent with prior theorizing of attentional monitoring as a waxing and waning process, rather than a process applied consistently throughout the ongoing task (e.g., Marsh, Hicks, & Cook, 2005; Scullin et al., 2013). A counter expectation might have been that, if people were monitoring for PM targets consistently in the form of a target checking process on the majority of trials (Guynn, 2003), they would be ready to respond by moving their mouse directly to the PM response field once a target appears. Our instructions reminding people to initiate mouse movement soon after the trial began, which is typical in studies measuring mouse movement data (e.g., Freeman & Ambady, 2010; Papesch & Goldinger, 2012), may have even restricted this possible behavioral outcome of relatively more direct trials for nonfocal conditions. An interesting methodological change would be to relax even further the restriction on mouse initiation to see if people will naturally be more likely to delay such movement in nonfocal conditions and travel directly to the PM field. In this case, one might expect main effects of both trajectory type (direct vs. reversal) and focality on initiation times.

The issues of initiation time and proportion of reversal trials brings up an important caveat to our consideration of monitoring, relative to the static trial starting method we adopted. Recent work by Scherbaum and Kieslich (2018) suggests that some cognitive paradigms might benefit from a dynamic starting methodology.⁵ Dynamic starting methods require participants to begin mouse movement *before* a stimulus appears, whereas static methods allow participants to initiate movement after the stimulus has appeared. Scherbaum and Kieslich demonstrated in a Simon effect paradigm that mouse trajectories were generally more consistent across trials with a dynamic start as opposed to a static start. Primary behavioral outcomes, such as the Simon effect itself, were significant in both starting methods. Nonetheless, mouse trajectory dynamic effects were smaller and more temporally compressed in the static starting method. Because the starting procedure could influence outcomes, one might question what results we might have obtained given a dynamic starting method. Given the stimulus and response area features shown in Fig. 1, one key difference with a dynamic starting method is that the initial mouse trajectories would overwhelmingly be toward the word/nonword response areas, thereby creating almost 100% reversal-like PM responses and few or no direct responses. Thus, it would likely exacerbate the overall character of the reversal trajectories shown in Fig. 2 (e.g., an even greater AUC and time to MD). As such, we speculate that our primary outcomes associated with reversal trajectories would not change. In other words, the PM RT would be longer for

nonfocal reversals, with a greater AUC, time to MD, and so forth, than for focal reversals. However, this does mean that our interpretation of focality differences in direct PM trials should be made with great caution.

The other obvious difference with a dynamic starting method is that initiation time becomes unavailable as a potential factor of interest. As we discussed in the introduction, we had the expectation that initiation times might be influenced by whether people were engaging in a relatively consistent monitoring strategy, and that such a strategy might prompt people to slow down their initiation time in order to consider the PM response alternative as a stimulus appeared. Such an outcome might also be consistent with one prediction of the delay theory of PM (Heathcote et al., 2015): namely, that people tend to slow down their ongoing-task responses when holding an intention. How and when during an entire trial this slowing occurs is uncertain when using a keypress methodology. Recall also that our instructions reminded participants to initiate their responses as quickly as possible on each trial and that we retained trials for analysis that took no more than 500 ms to initiate. Thus, in practice our paradigm likely was an amalgam of a static and a dynamic starting procedure. Of course, whether a dynamic procedure would change the character of our reported findings is an interesting empirical endeavor to be explored.

Another challenge for future work will be to leverage real-time response dynamics as a means to test theories of PM. As we suggested earlier, a useful measure from mouse-tracking paradigms with a static starting method is initiation time. Isolating this information from the overall RT directly is nearly impossible with a keypress methodology. On the one hand, encouraging people to move the mouse soon upon stimulus presentation—and certainly with a dynamic starting method—allows for a more valid interpretation of other aspects of the response trajectory (e.g., RT, AUC, and time to MD). On the other hand, giving people flexibility to begin mouse movements on their own schedule might reveal responses better correlated with the flexible use of attentional monitoring processes. One might imagine that a very controlled, target-checking approach (Guynn, 2003) for every stimulus would promote a slower initiation time and a higher ratio of direct-to-reversal movements to a PM response field in our paradigm. Regardless, the mouse-tracking approach has obvious empirical benefits over traditional keypress paradigms. In this sense it should prove to be a good companion measure to modeling of keypress RT distributions, because of direct access to real-time, dense sampling of behavior over the course of a trial response, as opposed to only the distribution of averaged responses.

In conclusion, we demonstrated patterns of retrieval dynamics suggesting that spontaneous retrieval occurs on PM trials, when people drift toward the typical ongoing-task decision response but then appropriately reverse toward the PM

⁵ We are grateful to Stefan Scherbaum for raising this important issue.

response field. In contrast, we argue that trials on which people are already prepared for the PM response, such as might be true when people are engaged in attentional monitoring, produced more direct mouse movements toward the PM response field. Unexpectedly, both types of response trajectories occurred in both the focal and nonfocal instruction conditions, indicating that such instruction conditions may not create consistent monitoring or consistent spontaneous-retrieval contexts over a block of trials. This pattern may also reveal evidence consistent with the dynamic-multiprocess view of PM (Scullin et al., 2013), that people change their monitoring process over the course of an ongoing task (Hicks et al., 2017; Rummel & Meiser, 2013). Using the mouse-tracking methodology to test the assumptions of PM-related constructs (e.g., focal vs. nonfocal conditions), the microstructure of PM retrieval processes (Marsh et al., 2003), and the predictions of PM theories should prove fruitful.

Appendix 1: Data related to Tables 1 and 2 (Tables 3 and 4, respectively), with all participants and all trials retained (i.e., no exclusions applied)

Table 3 Average proportions of correct PM responses across focality and task emphasis conditions

Emphasis	Focality							
	Focal			Nonfocal			Average	
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Standard	37	.95	.18	37	.68	.39	.82	.33
PM	38	.98	.05	37	.85	.20	.92	.16
LDT	38	.82	.37	41	.70	.40	.76	.39
Average		.92	.25		.74	.35		

Table 4 Simple main effects of focality for direct and reversal trajectory types, for various mouse-tracking measures and overall response time on correct PM trials

	Focal		Nonfocal		Row simple effect	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>p</i>	Partial η^2
Response time (RT)						
Direct	991	132	1,015	150	.33	.01
Reversal	1,143	110	1,249	154	< .01	.15
Area under curve (AUC)						
Direct	0.06	0.07	0.07	0.06	.79	< .01
Reversal	0.81	0.45	1.08	0.39	< .01	.10
Maximum deviation (MD)						
Direct	0.09	0.11	0.10	0.11	.83	< .01
Reversal	0.62	0.28	0.80	0.19	< .01	.12
Time to MD						
Direct	540	93	592	124	< .01	.06
Reversal	544	68	645	84	< .01	.31

All measures are averaged medians (analyses of participant means showed similar results); RT and Time to MD are in milliseconds; the two-way interaction between focality and trajectory type was significant for each of these measures; *p* values are not corrected for family-wise Type I error inflation in this table

Appendix 2

Table 5 Various descriptive statistics for non-PM word stimuli accurately called “word” in the lexical decision task

	Focal			Nonfocal		
	Block 2	Block 3	Block 3 Adj.	Block 2	Block 3	Block 3 Adj.
Initiation time						
Direct	282 (104)	278 (103)	280 (78)	2,710 (90)	261 (114)	276 (78)
Reversal	271 (107)	247 (117)	241 (100)	249 (105)	246 (112)	252 (101)
Response time (RT)						
Direct*	957 (82)	951 (86)	951 (58)	943 (95)	970 (94)	980 (58)
Reversal*	1,223 (172)	1,214 (167)	1,197 (154)	1,169 (118)	1,235 (181)	1,251 (154)
Area under curve (AUC)						
Direct	0.02 (0.04)	0.03 (0.05)	0.02 (0.04)	0.02 (0.04)	0.03 (0.04)	0.02 (0.04)
Reversal	0.31 (0.19)	0.34 (0.18)	0.35 (0.12)	0.34 (0.12)	0.34 (0.20)	0.34 (0.11)
Maximum deviation (MD)						
Direct	0.04 (0.08)	0.05 (0.10)	0.04 (0.10)	0.04 (0.08)	0.05 (0.08)	0.05 (0.08)
Reversal	0.61 (0.25)	0.68 (0.20)	0.69 (0.16)	0.65 (0.18)	0.66 (0.24)	0.66 (0.16)
Time to MD						
Direct*	509 (55)	511 (57)	515 (42)	507 (68)	530 (63)	534 (42)
Reversal*	616 (135)	606 (90)	602 (99)	596 (83)	639 (117)	639 (99)

Averages of medians in all cells with *SD* in parentheses; the “Adj.” column indicates Block 3 medians adjusted by a Block 2 covariate. * Significant difference between focal versus nonfocal comparison of covariate-adjusted Block 3 means—these are bolded values in a given row of the table; except for initiation time, we found a main effect of trajectory type for all other measures, such that reversal data were significantly larger than direct data

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

- Abney, D. H., McBride, D. M., Conte, A. M., & Vinson, D. W. (2015). Response dynamics in prospective memory. *Psychonomic Bulletin & Review*, 22, 1020–1028.
- Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, 39, 445–459.
- Einstein, G. O., & McDaniel, M. A. (1990). Normal aging and prospective memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 717–726. <https://doi.org/10.1037/0278-7393.16.4.717>
- Einstein, G. O., & McDaniel, M. A. (2005). Prospective memory: Multiple retrieval processes. *Current Directions in Psychological Science*, 14, 286–290. <https://doi.org/10.1111/j.0963-7214.2005.00382.x>
- Freeman, J. B., & Ambady, N. (2010). MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods*, 42, 226–241. <https://doi.org/10.3758/BRM.42.1.226>
- Graf, P. (2005). Prospective memory retrieval revisited. In N. Ohta, C. M. MacLeod, & B. Uttil (Eds.), *Dynamic cognitive processes* (pp. 305–332). New York, NY: Springer.
- Gunn, M. J. (2003). A two-process model of strategic monitoring in event-based prospective memory: Activation/retrieval mode and checking. *International Journal of Psychology*, 38, 245–256. <https://doi.org/10.1080/00207590344000178>
- Heathcote, A., Loft, S., & Remington, R. W. (2015). Slow down and remember to remember! A delay theory of prospective memory costs. *Psychological Review*, 122, 376–410. <https://doi.org/10.1037/a0038952>
- Helman, E., Stoller, R. M., & Freeman, J. B. (2014). Advanced mouse-tracking analytic techniques for enhancing psychological science. *Group Processes & Intergroup Relations*, 18, 384–401. <https://doi.org/10.1177/1368430214538325>
- Hicks, J. L., Franks, B. A., & Spitzer, S. N. (2017). Prior task experience and comparable stimulus exposure nullify focal and nonfocal prospective memory retrieval differences. *Quarterly Journal of Experimental Psychology*, 70, 1997–2006. <https://doi.org/10.1080/17470218.2016.1217891>
- Horn, S. S., & Bayen, U. J. (2015). Modeling criterion shifts and target checking in prospective memory monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41, 95–117. <https://doi.org/10.1037/a0037676>
- Johnson, A., Mulder, B., Sijbinga, A., & Hulsebos, L. (2012). Action as a window to perception: Measuring attention with mouse movements. *Applied Cognitive Psychology*, 26, 802–809. <https://doi.org/10.1002/acp.2862>
- Kieslich, P. J., & Henninger, F. (2017). Mousetrap: An integrated, open-source mouse-tracking package. *Behavior Research Methods*, 49, 1652–1667. <https://doi.org/10.3758/s13428-017-0900-z>
- Kliegel, M., Martin, M., McDaniel, M. A., & Einstein, G. O. (2001). Varying the importance of a prospective memory task: Differential effects across time- and event-based prospective memory. *Memory*, 9, 1–11. <https://doi.org/10.1080/09658210042000003>
- Kliegel, M., Martin, M., McDaniel, M. A., & Einstein, G. O. (2004). Importance effects on performance in event-based prospective

- memory tasks. *Memory*, 12, 553–561. <https://doi.org/10.1080/09658210344000099>
- Kurtz, M. (2017). *Investing the processes underlying cost to the ongoing task and after effects in prospective memory*. Unpublished master's thesis, Technische Universität Dresden.
- Magnuson, J. S. (2005). Moving hand reveals dynamics of thought. *Proceedings of the National Academy of Sciences*, 102, 9995–9996. <https://doi.org/10.1073/pnas.0504413102>
- Marsh, R. L., Hicks, J. L., & Cook, G. I. (2005). On the relationship between effort toward an ongoing task and cue detection in event-based prospective memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 68–75. <https://doi.org/10.1037/0278-7393.31.1.68>
- Marsh, R. L., Hicks, J. L., Cook, G. I., Hansen, J. S., & Pallos, A. L. (2003). Interference to ongoing activities covaries with the characteristics of an event-based intention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 861–870. <https://doi.org/10.1037/0278-7393.29.5.861>
- Marsh, R. L., Hicks, J. L., & Watson, V. (2002). The dynamics of intention retrieval and coordination of action in event-based prospective memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 652–659. <https://doi.org/10.1037/0278-7393.28.4.652>
- McDaniel, M. A., & Einstein, G. O. (2000). Strategic and automatic processes in prospective memory retrieval: A multiprocess framework. *Applied Cognitive Psychology*, 14, S127–S144. <https://doi.org/10.1002/acp.775>
- Meier, B., Zimmermann, T. D., & Perrig, W. J. (2006). Retrieval experience in prospective memory: Strategic monitoring and spontaneous retrieval. *Memory*, 14, 872–889. <https://doi.org/10.1080/09658210600783774>
- Papesh, M. H., & Goldinger, S. D. (2012). Memory in motion: Movement dynamics reveal memory strength. *Psychonomic Bulletin & Review*, 19, 906–913. <https://doi.org/10.3758/s13423-012-028103>
- Rummel, J., & Meiser, T. (2013). The role of metacognition in prospective memory: Anticipated task demands influence attention allocation strategies. *Consciousness and Cognition*, 22, 931–943. <https://doi.org/10.1016/j.concog.2013.06.006>
- Rummel, J., Wesslein, A. K., & Meiser, T. (2017). The role of action coordination for prospective memory: Task-interruption demands affect intention realization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43, 717–735. <https://doi.org/10.1037/xlm0000334>
- Scherbaum, S., & Kieslich, P. J. (2018). Stuck at the starting line: How the starting procedure influences mouse-tracking data. *Behavior Research Methods*, 50, 2097–2110. <https://doi.org/10.3758/s13428-017-0977-4>
- Scullin, M. K., McDaniel, M. A., & Shelton, J. (2013). The dynamic multiprocess framework: Evidence from prospective memory with contextual variability. *Cognitive Psychology*, 67, 55–71. <https://doi.org/10.1016/j.cogpsych.2013.07.001>
- Smith, R. E. (2003). The cost of remembering to remember in event-based prospective memory: Investigating the capacity demands of delayed intention performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 347–361. <https://doi.org/10.1037/0278-7393.30.4.756>
- Spivey, M. J. (2007). *The continuity of mind*. New York, NY: Oxford University Press.
- Spivey, M. J., & Dale, R. (2006). Continuous dynamics in real-time cognition. *Current Directions in Psychological Science*, 15, 207–211. <https://doi.org/10.1111/j.1467-8721.2006.00437.x>
- Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences*, 102, 10393–10398. <https://doi.org/10.1073/pnas.0503903102>
- Uttl, B., White, C. A., Gonzalez, D. W., McDouall, J., & Leonard, C. A. (2013). Prospective memory, personality, and individual differences. *Frontiers in Psychology*, 4, 130:1–15. <https://doi.org/10.3389/fpsyg.2013.00130>