

©American Psychological Association, 2022. This paper is not the copy of record and may not exactly replicate the authoritative document published in the APA journal. The final article is available, upon publication, at: [DOI TBA]

When did this happen?

Indicators of accuracy for dating recent and remote personal events

Aleš Neusar^a and Eva Rubínová^b

^a Department of Psychology, University of Ostrava, Reální 5, 702 00 Ostrava, Czech Republic. ORCID: <https://orcid.org/0000-0001-9574-7761>

^bDepartment of Psychology, Thompson Rivers University, 805 TRU Way, Kamloops, British Columbia, V2C 0C8, Canada. ORCID: <https://orcid.org/0000-0001-9790-6180>

Author Note

Correspondence concerning this article should be addressed to Eva Rubínová, who is now at School of Psychology, University of Aberdeen. E-mail: eva.rubinova@abdn.ac.uk. All data will be made publicly available at the Open Science Framework upon publication and can be accessed at <https://osf.io/u594x/>.

Abstract

Memory for time is often imperfect. We investigated indicators of dating accuracy for recent and remote personal events. In Study 1, 78 participants dated approximately 20 events from the past three to eleven weeks, and in Study 2, 40 participants dated approximately 25 events three to five years old. For date verification, events were obtained from participants' partners, who kept a diary (Study 1) or retrieved events from personal records (Study 2). In both studies, we found that confidence, dating strategy indicating known dates, and a direct connection with temporal landmarks were associated with higher dating accuracy. High importance, direct experience, and events embedded within extended periods indicated higher dating accuracy for recent events; high vividness was associated with higher dating accuracy for remote events. Our results suggest that confidence, dating strategy, and phenomenological characteristics can provide useful indication of dating accuracy.

Key words: autobiographical memory, dating accuracy, temporal schema, temporal landmark, confidence-accuracy

General Audience Summary

When did you have your last vaccination? The answer to temporal questions requires a series of reconstructive processes involving the recall of event details, context, and other characteristics that may help with the estimate. The exception are events that serve as “landmarks” in people’s temporal memory. We investigated indicators of dating accuracy for personal events that happened in the past three to eleven weeks (recent events, Study 1) and personal events that happened three to five years ago (remote events, Study 2). To evaluate dating accuracy, we recruited couples: one member (secretly) provided descriptions and dates of events of their partner’s personal events (collected via a diary in Study 1 and from personal records in Study 2); the other member was then interviewed about the events. In both studies, we asked participants to think aloud when trying to estimate when each event happened and evaluate various event characteristics. We then coded whether the participant was the central character in the event or whether another person was central, whether the event happened within one day or was extended, and whether an association with a temporal landmark event served as the main source for the final date estimate. In both studies, we found that events were dated more accurately when participants were confident in their estimates, their dating strategy indicated known dates, and when date was reconstructed via a direct association with a temporal landmark. High event importance, direct experience, and events embedded within extended periods indicated higher dating accuracy for recent events; high vividness (i.e., detailed memory) was associated with higher dating accuracy for remote events. It seems that confidence, dating strategy, and event characteristics can provide indication of dating accuracy that may be useful in interviewing settings where date verification is not available.

When did this happen? Indicators of accuracy for dating recent and remote personal events

Medical professionals, forensic interviewers, or sociologists frequently collect information about personal events. Temporal estimates may serve as a base for decisions regarding treatment (e.g., Means et al., 1988; Schwarz, 2007) or establish the start of offences (e.g., Fisher & Geiselman, 2010; Yoshihama et al., 2002). In these applied contexts, the accuracy of temporal estimates becomes crucial. However, memory for time is often imperfect (e.g., Wagenaar, 1986). Our aim in the current research was to investigate dating strategy, personal event characteristics, and confidence ratings as potential indicators of dating accuracy, and the consistency of these indicators across recent (i.e., three to eleven weeks old) and remote (i.e., three to five years old) personal events. We also tested the effectiveness of a simple calendar method in aiding temporal estimation.

Temporal attributes are scarcely directly associated with the event (Friedman, 1993) and, instead, must be reconstructed (e.g., Ben Malek et al., 2017; Janssen et al., 2006; Thompson et al., 1993; 1996). People use their knowledge about when events happen—temporal schemata—to constrain the search to specific units of time (e.g., dinner starts at 7PM, language class happens on Thursday; Larsen et al., 1995). To arrive at a more precise estimate (i.e., which Thursday in a month), people may use temporal landmarks—events that provide boundaries (e.g., the beginning of the language course; Brown, 2016; Lee & Brown, 2004; Shum, 1998). Associations with other events, retrieval of further event details, or specific characteristics of the event may then provide further guidance on the final temporal estimate. Importantly, the processes that accompany the reconstruction of temporal information seem to be universal for temporal scales ranging from minutes through days to months and years. But how much can we trust temporal estimates?

When dates are known, the accuracy tends to be high. Eighty percent of known date estimates were accurate (Betz & Skowronski, 1997) or were associated with the smallest dating error (Thompson et al., 1993). However, dates are known only for a minority of events (i.e., less than 20% of self-experienced events; Betz & Skowronski, 1997; Thompson et al., 1993). Temporal landmarks form a specific category of events for which the dates are typically known and highly accurate (e.g., Gaskell et al., 2000; Loftus & Marburger, 1983). Consequently, dates of personal events that are associated with temporal landmarks (e.g., a language class following return from holiday) are more accurate than dates of events reconstructed with the use of other strategies (e.g., estimating of the number of events that occurred since the dated event; Smith & Thomas, 2003; Thompson et al., 1993).

People may also rely on their temporal knowledge. Temporal schemata or other forms of knowledge associated with the regularity of life in a given place aid the reconstruction of temporal information, but only to a limited degree (e.g., only in certain time units). That is, the retrieval of personal events may cue general knowledge (e.g., when reconstructing the time one met with a friend, they may recall passing by a café that was closed, suggesting that the meeting likely happened after 5PM) or personal knowledge about time (e.g., when reconstructing what day one received an uninvited call, they may recall that it was on the way to a regular Wednesday volleyball practice; e.g., Larsen et al., 1995; Thompson et al., 1993). In both cases, temporal schemata anchor the time in terms of a specific unit (e.g., hour or day of the week), but provide little guidance for other units of time.

The success of the reconstruction of temporal information may also depend on how well an event is remembered. It is likely that highly memorable events would provide cues leading to more accurate temporal estimates. But what makes an event memorable? Event recency, i.e., the time between the present and the occurrence of the event, can play a role due to reduced retroactive interference. Therefore, we could expect better memory and higher

dating accuracy for more recent events (e.g., Bradburn, 2000). Direct experience (i.e., self-events versus other-events) should also contribute to higher memorability, and is typically associated with higher dating accuracy (Betz & Skowronski, 1997; Skowronski et al., 1994). Events embedded within longer periods (e.g., returning from holiday) may be remembered better and may become temporal landmarks more frequently than standalone events because they provide period boundaries (Brown, 2016; Shum, 1998). Next, rehearsal or social sharing may also contribute to memorability (e.g., Neisser, 1988; Sutin & Robins, 2007; Williams et al., 2008) and potentially also to higher dating accuracy.

Memorability can be also indirectly measured via subjective phenomenological characteristics of events, such as importance, vividness, or emotional intensity (e.g., Sutin & Robins, 2007; Rubin et al., 2003; Wade & Adams, 1990). People remember important events in greater detail than trivial events (Ley, 1972; Ritchie et al., 2006), although importance may be re-evaluated depending on other life experiences and changes that come along with them (White, 2020; Williams et al., 2008). The amount of recalled details (i.e., vividness), uniqueness, emotionality, or coherence of the event story may be related to importance and memorability (Ritchie et al., 2006; Rubin & Kozin, 1984; Rubin et al., 2003; White, 2020). In addition to event characteristics, participants can also provide metacognitive evaluations of their performance that may serve as an indication of accuracy.

In the eyewitness identification literature, researchers have found that confidence ratings may serve as reliable predictors of accuracy of decisions if confidence ratings were collected during the initial memory test immediately after the memory-based decision and in the absence of interviewer feedback (Wixted & Wells, 2017). The relationship between confidence and accuracy is particularly strong at the higher levels of confidence (Saraiva et al., 2020; Wixted & Wells, 2017). Finally, researchers have investigated various methods that may aid the accuracy of date estimates.

Aids for the retrieval of temporal information typically involve a calendar—a graphical display of a timeline that may include public events and that may be divided into units of time that are of interest for a given interview (see Belli, 1998; Belli & Callegaro, 2009; Glasner & Van der Vaart, 2009; Roberts & Horney, 2010). During or before an interview, participants are typically asked to populate the calendar with their personal landmark events, which serve as cues for later retrieval of personal events (e.g., Belli, 1998). Calendar methods have been found to improve recall accuracy (Belli & Callegaro, 2009; Glasner & Van der Vaart, 2009; Van der Vaart & Glasner, 2011).

Current Research

In two studies, we examined indicators of dating accuracy for personal events, and the effectiveness of a calendar method in improving dating accuracy. In both studies, participants were provided with event descriptions and asked to estimate when the events happened while thinking aloud and subsequently rated several event characteristics and confidence. Event descriptions including temporal information were acquired from participants' proxies (i.e., partners). To ensure verifiable temporal information, in Study 1, proxies kept diaries of participants' personal events for a period of 6 weeks; in Study 2, proxies retrieved event temporal information from their personal records.

The recruitment of couples for this study is a methodological advancement that overcomes potential effects associated with studies in which participants themselves kept diaries and were later interviewed about their personal events (e.g., Linton, 2000; Skowronski et al., 1991; Wagenaar, 1996; White, 2020). In other words, as an unexpected memory test following incidental learning, our method had higher ecological validity and enabled us to make comparisons with previous research using diary methods. Next, diary studies examining event characteristics as indicators of dating accuracy typically focus on personal events several weeks/months old (e.g., Thompson et al., 1996). Our main motivation for examining

dating accuracy for both recent as well as remote events was the ability to identify general indicators of dating accuracy (i.e., indicators relatively independent of time scale), respectively to identify the time scale limitations of various indicators.

We expected that dating strategy that involved direct retrieval of the temporal estimate or connection with a landmark event would be associated with higher dating accuracy. Next, we expected that event characteristics that likely contribute to higher memorability, such as recency, direct experience, events embedded within extended periods, and frequent sharing, as well as subjective characteristics, such as importance, vividness, uniqueness, or emotional intensity would be associated with higher dating accuracy. More frequently occurring and regular events are typically associated with temporal schemata; therefore, we expected that participants would be more accurate particularly in their temporal estimates in the units of the schema (i.e., days of the week for recent and months for remote events). We also expected that participants' confidence would be indicative of accuracy. Finally, we expected that using the calendar aid would benefit dating accuracy.

Study 1: Recent Events

Method

Design

This study was a 2-group (interview: calendar/no calendar) between-subjects design. The dependent variables were (1) dating accuracy (yes/no) and (2) accuracy of the day of the week estimate (yes/no).

Participants

Seventy-eight couples (i.e., participants and their proxies) aged 24 to 46 years ($M = 31.4$, $SD = 4.3$; all White ethnicity) were recruited using flyers at public places and snowball sampling, where interviewers (A. N. and E. R.) and three trained research assistants sent out emails initiating the snowballs and participants then further distributed flyers among their

acquaintances (details regarding interviewer training are provided as [supplemental materials](#) on the Open Science Framework [OSF]). The flyers advertised a study in memory for everyday events for which the researchers sought to recruit couples. Interested couples were informed that no sensitive information would be collected from either partner. Proxies were, in addition, informed that the study focused on dating personal events, but they were asked to not reveal this information to their partners.

The inclusion criteria specified that: (1) the couples needed to live together for at least half a year before the interview ($M = 7.6$, $SD = 4.5$ years), (2) needed to be present in the Czech Republic (the country where the interview took place) without a break longer than two weeks during the reference period (March 14 to June 5, 2011; note that this period precedes and extends the diary period mentioned below to provide space for dating errors), (3) had to live together during the reference period, and (4) needed to be in regular contact when either partner was away from home for more than two days. The role of participants (35 self-identified as females and 43 as males; 11 completed high school and 67 had university education) was to attend the study interview; the role of proxies (43 self-identified as females and 35 as males) was to keep online diaries of their partners' personal events (see Procedure for further details). Couples were compensated with a small present worth approximately \$6 for their participation.

Materials and Procedure

Diary Entries. Proxies (secretly) kept online diaries for a period of six weeks (Monday April 11 to Sunday May 29, 2011; one proxy supplied an event that occurred on March 23, i.e., before diary period, and we kept the event in the study). Their task was to make several entries per week, recording events experienced by their partners, the couple, or the whole family. They were instructed to record any noteworthy events and skip days for which they had little information about what happened to their partners. For each event,

proxies were asked to: (1) provide a description that would be detailed enough so that the event could be distinguishable from other events; (2) the day of the week and (3) the exact date of the event; (4) evaluate the emotional intensity (low/moderate/high) and (5) emotional valence of the event for the participant (negative/ambivalent/positive), or answer “I don’t know” if proxies felt that they did not have enough information to evaluate the emotional characteristics of the event.

Selection of Personal Events. Before the participant interview, proxies were provided with all their diary entries and asked to select 22 events based on the following criteria: (1) avoid events with obvious dates (e.g., family member’s birthday); (2) avoid sensitive events or, if selected, delete any sensitive details from the event description; (3) avoid events that occurred on one day or that were strongly related (e.g., “travelling to a cottage” and “having dinner at the cottage after arrival”); (4) delete any temporal cues or hints from the event description (e.g., “Tuesday tennis training with a new racket”); (5) select events from the full range of the period. A. N. checked that all selected events complied with the criteria; if needed, A. N. asked the proxy for further adjustment or excluded the event. The final number of events presented for an interview ranged between 10 and 22 ($M = 20.0$, $SD = 2.7$, $Med = 21.0$) across participants.

After event selection, proxies were provided with written instructions and asked to provide further ratings for each event. We asked proxies how accurate they thought their partners (interviewed participants) would be in their estimates of (1) the day of the week (accurate/inaccurate) and (2) the week (accurate/ \pm one week error/ \pm two weeks error). Proxies also provided ratings of (3) importance of the event for their partners (low/moderate/high) and the level of (4) sharing of the event (not shared/few times (1–3 times)/many times). A description of the relationship between dating accuracy and event importance and confidence estimates and the correspondence between participant and proxy estimates are reported in

Neusar and Van der Vaart (2018). Four public events were added to the personal events (for further details, see [OSF](#)); these results were not of central interest in this study and are not reported here.

Pocket Calendar. All participants received a small “pocket calendar.” The calendar showed exact dates organized in a weekly fashion and in a typical calendar, and weekends and holidays were highlighted there (i.e., Easter Monday on April 25, and Labor Day on Sunday May 1, and Victory in Europe Day on Sunday May 8). In order to limit boundary effects in dating personal events, we extended the interview period: we added 4 weeks before and two weeks after the six-week diary period during which personal events were recorded.

Enhanced Calendar and No Enhanced Calendar Conditions. Participants in the enhanced calendar condition received a second large calendar, which provided space for a visual organization of events. Participants additionally generated temporal landmarks before the dating task (see below). Participants in the no enhanced calendar condition received a list of event descriptions on a sheet of paper and were not asked to generate temporal landmarks before the dating task. For brevity, we will refer to the “no enhanced calendar condition” as the “no calendar condition”.

Interview. All interviews took place between June 6 and June 29 (M and Med = June 13), that is between 8 and 31 days (M = 15 days) after the diary period. Interviews were audio recorded and conducted in participants’ homes, at the university, or in a quiet café. There were 4 phases of the interview depending on the condition: (1) demographics, self-evaluation, and dating instructions (followed by landmark recall for participants in the calendar condition); (2) event dating; (3) event characteristics and confidence ratings; and (4) demographic follow-up, feedback, and debriefing.

Demographics, Self-Evaluation, and Dating Instructions. Interviewers first checked if participants learned any detailed study information from their proxies and asked about

participants' expectations (no participant described any hypotheses-relevant information). Then, participants responded to a series of demographic questions, including their birthday, education, length of relationship with their partner, any children (names were required for the evaluation of content of some personal events), and self-evaluation of memory (i.e., open questions regarding memory for content and dating of events from the past few months followed by 3-point relational scale: better than, similarly as, or worse than other people I know; and 4-point scale questions evaluating the difficulty to estimate the day of the week or exact date of personal events: very difficult, rather difficult, rather easy, very easy). Finally, the interviewer presented participants with instructions for the dating task, which differed for the calendar and no calendar conditions.

In the calendar condition, the interviewer showed the participants both calendars, and explained that they cover the reference period and highlight public holidays. Participants were then asked to think about events from that period, especially events that they remembered well and knew when they happened, write labels for each of these events on a sticky note and paste the note onto the large calendar. Participants were told that for any events that extended over one day, they could mark the beginning and the end, or they could split the event into several day events. Participants could change the positions (i.e., dates) of these landmark events. In the no calendar condition, the interviewer only introduced the small calendar (i.e., participants did not recall landmark events).

Event Dating. Selected personal and public events were randomly ordered and one-by-one presented to participants. In the calendar condition, the interviewer read out and handed the participant the event descriptions printed on sticky notes; in the no calendar conditions, the interviewer read out the first event description that was printed on the sheet. Participants were asked to think aloud about when each of these events happened. In the calendar condition participants pasted events into the large calendar; in the no calendar

condition, participants wrote the exact date in the space provided. During this task, the interviewer coded whether the date was known (i.e., immediately provided) or reconstructed (i.e., participants needed to recall additional information before they arrived at the date); if unclear, the interviewer asked whether the participant knew the date or whether they needed to consider other information before providing the estimate. In the calendar condition, the interviewer marked any dated events that were also temporal landmarks. For participants who took long or seemed like they were struggling with the temporal estimation, the interviewers attempted to facilitate temporal estimation by asking about general indicators that might provide dating constraints. For example, interviewers could ask participants to think about whether the event occurred during the working week or the weekend, whether the event was associated with another event they could remember, whether they thought the event occurred relatively recently or further in the past, or what was the weather like on the day.

Immediately after providing a temporal estimate, the interviewer assessed: (1) frequency of the event by asking the participant whether another event similar to the dated one occurred more frequently during the reference period (3-point scale: event happened once, 2–3 times, more times), and (2) event regularity by asking whether the event regularly occurred on a specific day (yes/no, if yes, participants were asked to specify the regularity). In the calendar condition, the interviewer asked whether the event was one of the landmark events.

Event Characteristics and Confidence Ratings. After participants dated all events, they were presented with each event one-by-one and were asked to evaluate the degree of personal importance (low/moderate/high) and sharing of each event (not shared/shared few times (1–3 times), many times, not shared). Participants were then asked a series of confidence ratings: (1) “Do you think that you estimated the day of the week accurately?” (no/yes); (2) “Do you think that you estimated the week accurately?” (no/yes; if no, participants were asked to estimate how many weeks prior or after the estimate the event

could have happened; see also Thompson et al., 1996); (3) participants who answered “yes” to the accuracy of both day of the week and week estimates were asked: “Would you be willing to testify under oath for the date?” (no/yes).

Demographic Follow-Up, Feedback, and Debriefing. Participants were then asked whether they had any of the public events associated with a personal event. Participants in the calendar condition were asked whether they thought that the calendar helped them in the dating task, and whether it helped specifically with the accuracy or speed of dating (separate ratings; 3-point scale: yes, helped a lot; helped a little; no). All participants were asked whether any of the public holidays highlighted in the calendar helped them in the dating task (and, if yes, which ones). Participants then evaluated the temporal regularity of their life (irregular/rather irregular/rather regular/very regular) and responded whether they had any regular free-time activities and whether their job involved regularity (no/yes). Interviewers then asked for feedback on the interview, evaluated interviewee and interviewer tiredness and motivation (low/moderate/high). Note that these ratings were not of central interest in this paper and are not reported here—interested readers may use the data provided on the [OSF](#).

Coding

Based on the event descriptions and audio recordings of the dating protocols, E. R. coded events as: (1) self-events or other-events, (2) single-day or extended events, and (3) events associated with temporal landmarks. Self-events were defined as events in which the participant was the central character (e.g., a work meeting in a different city) and events in which someone else was the central character but that created a unique event for the participant (child-related events, such as a child’s illness, were typically coded as “self”). Other-events were unique for another person who was the central character, and there was little or unclear involvement of the participant (e.g., when the participant’s partner left to play Dungeons and Dragons”), or participant’s involvement was not unique (e.g., when the

participant picked up his partner and child from their first roller-skating trip, where the pickup that involved the participants was one of many pickups, although the event was unique for the partner).

The Dungeons and Dragons event also serves as an example of an extended event—the dated event in fact marked the beginning of an event that extended over more than one day. Other examples of unique events that occurred within an extended period were an opening night at the participant's art exhibition that involved spending several days in a different city, and a friend's visit of the participant when he was on vacation with his family. These events were coded as extended events but were associated with a single date depending on the event description (e.g., the beginning of a long weekend; the opening of the exhibition; the day the friend visited during vacation). Single-day events occurred on one day.

For temporal landmark connection, events were categorized as having a direct connection with a landmark in cases when the landmark had a causal relationship to the event or when the landmark event was the main source of the date estimate; other events were coded as not having a direct landmark connection. Landmarks were personal (e.g., leaving for a conference) or public (e.g., Easter or public holidays). Temporal landmark connection was coded for events in both conditions (with and without the enhanced calendar) because this measure was coded based on dating protocols.

To obtain estimates of inter-rater reliability, a trained research assistant independently coded complete data from 12 randomly selected participants (~15% of the sample; details of coder training can be found on the [OSE](#)). The standard for computing estimates of inter-rater reliability is Cohen's kappa due to its control for chance agreement (Cohen, 1960). However, Cohen's kappa results in significantly reduced estimates of inter-rater reliability in case of low base rates (e.g., Xu & Lorber, 2014). The majority of events in our data were self-events, single-events, and events that had no direct association with a temporal landmark, resulting in

low base rates for these measures (Table 2). To prevent underestimating inter-rater reliability, we opted for computing agreement rates rather than Cohen's kappa. Agreement was high: 96% for self/other events, 88% for extended/single events, and 83% for landmark association (no/direct). E. R.'s codes were used for all statistical analyses.

E. R. further categorized events into content categories (e.g., sports, shopping, health, or relationship events) but these categories were not of primary interest in this paper and are not reported here (see Literáková & Neusar, 2011).

Measures

We measured fourteen variables. (1) Recency was the number of days between the day the event occurred and the interview. (2) Dating strategy (date known/reconstructed) was an interviewer's evaluation of participant dating based on whether the participant immediately provided the final date (known) or whether the participant retrieved additional information before providing the final date estimate (reconstructed). For participants in the calendar condition, the interviewer marked (3) dated events that were also temporal landmarks that participants retrieved before the dating task (no/yes).

Participants evaluated (4) the frequency of the event within the reference period (event happened once/2–3 times/more times; categorical; note that we treated all ordinal 3-point scales as categorical) and (5) whether event regularly occurred on a specific day (no/yes). Participants also evaluated (6) event importance (low/moderate/high) and (7) frequency of sharing (not shared/few times (1–3 times)/many times).

For the evaluation of confidence in participants' temporal estimate, participants first reported whether they were confident in their day-of-the-week and week estimates (separate ratings; no/yes). If participants indicated the lack of confidence in the week estimate, we asked how many weeks before and after the estimated week the event could have occurred (we did not use these data for any analyses, but they are available on the [OSF](#)). We created (8)

a confidence measure (no/yes) by assigning “yes” in cases where participants were confident in their day-of-the-week and week estimates and “no” in cases where participants indicated the lack of confidence in at least one estimate. Participants who expressed confidence in both estimates were asked if they were willing to (9) testify under oath for the accuracy of the date estimate (no/yes).

The next two measures were collected from proxies, who evaluated: (10) the intensity of emotional involvement of the participant in the event (low/moderate/high/I don’t know), and (11) the emotional valence of the event for the participant (negative/ambivalent/positive/I don’t know). Proxies provided these evaluations only when they felt they had sufficient information; “I don’t know” responses were treated as missing values.

The final three measures were coded by E. R. from event descriptions and dating protocols: (12) self/other events were coded based on participant centrality in the event; (13) single-day/extended events were coded based on whether they happened on one day or within an extended period. Finally, events were coded as (14) having a direct association with a temporal landmark (no/yes) if a temporal landmark served as a base of the final date estimate.

Statistical analyses

Due to the nature of the typical performance in dating tasks (i.e., for recent events, people are more accurate in estimating the day of the week than the absolute date and more frequently confuse adjacent days and days of the week; Larsen et al., 1995), dating accuracy data are typically not normally distributed (see Figure 1). These regularities present limits for the estimation of a linear relationship between factors of interest and dating error (signed or absolute); therefore, we treated accuracy as a binomial variable (accurate/inaccurate).

We analyzed data in a series of generalized linear mixed models (GLMMs) in R (R Core Team; 2020) using the `lme4` (Bates et al., 2015) and `lmerTest` packages (Kuznetsova et al., 2017). Because each participant provided date estimates for multiple events, the data

were not independent. To account for these dependencies, we used participants as a random intercepts factor in all models (Finch et al., 2014). All categorical fixed factors were coded with successive difference contrasts (Schad et al., 2020) from the MASS package (Venables & Ripley, 2002), so that each contrast compared two adjacent levels of a factor (e.g., for importance, the first contrast compared low versus moderate importance, and the second contrast compared moderate versus high importance).

To aid interpretation of the results, the regression coefficients were exponentiated to indicate the odds ratio (*ORs*) of the change in the outcome variable between levels of a given contrast (e.g., the increase in dating accuracy between events evaluated as low and moderately important). Ninety-five percent confidence intervals are presented in brackets. To limit Type I error, we computed a boundary *p*-value based on the 30 *p*-values from all tests examining the relationship between measured variables and dating accuracy (Benjamini & Hochberg, 1995); therefore, only *p*-values $\leq .010$ were considered statistically significant.

Packages `ggplot2` (Wickham, 2016) and `gridExtra` (Auguie, 2017) were used for visualizations. Data, coding functions, and analysis scripts are available on the [OSF](#).

Results

Patterns of Dating Error

The left panel of Figure 1 shows the patterns of dating error, including the typical cumulations of errors in multiples of 7 days indicating confusions of weeks. Dating errors were evenly distributed between backward telescoping (i.e., dating events as occurring in more distant past, 51%) and forward telescoping (i.e., dating events as occurring in more recent past, 49%).

Table 1 shows proportions of accurate estimates and proportions of confusion across days of the week. The confusion matrix replicates the typical pattern showing how temporal schemata guide estimation (see also Thompson et al., 1996): day of the week estimates were

most frequently accurate, and confusions most frequently occurred across adjacent days of the week. There also seems to be a more fine-grained division of the week into the working week and the weekend with Friday as a relatively unique day. Specifically, for events that happened between Monday and Thursday, confusions with adjacent days in the middle of the working week occurred more frequently than confusions with days outside this period. For events that happened on Saturday and Sunday, confusions with the other day of the weekend were more frequent than confusions with days outside of the weekend. For events that happened on Friday, confusions were approximately evenly distributed across other days of the week.

Figure 1

Distribution of dating errors (left) and absolute dating error and recency (right)

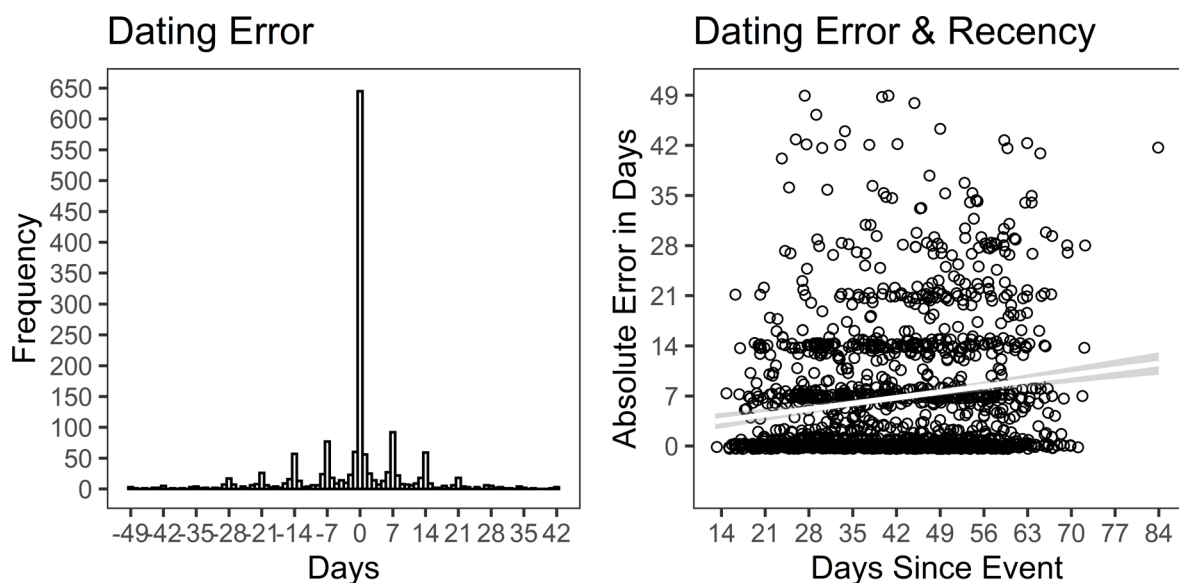


Table 1

Day of the week confusion matrix

DOW	N_{Events}	Estimated Day of the Week						
		Mon	Tue	Wed	Thu	Fri	Sat	Sun
Mon	176	.65	.14	.07	.07	.04	.02	.05
Tue	177	.14	.49	.11	.10	.04	.02	.02
Wed	203	.08	.19	.53	.15	.04	.02	.01
Thu	185	.03	.08	.17	.54	.05	.01	.01

Fri	236	.05	.06	.07	.09	.68	.05	.02
Sat	313	.02	.02	.02	.03	.10	.77	.11
Sun	269	.03	.03	.03	.03	.05	.12	.78

Note. DOW = day of the week.

We assessed the accuracy of day of the week estimates based on the day of the event. We coded days of the week with successive difference contrasts (i.e., we compared the odds of accurate estimates between Sunday and Monday; Monday and Tuesday, etc.). There were three significant contrasts (all with small effect sizes). Day of the week estimates were more accurate for events that happened on Sunday compared to Monday, $OR = 2.12 [1.32, 3.39]$, $z = 3.12$, $p = .002$, Monday compared to Tuesday, $OR = 1.85 [1.16, 2.95]$, $z = 2.58$, $p = .010$, and Friday compared to Thursday, $OR = 1.88 [1.25, 2.81]$, $z = 3.05$, $p = .002$. The other three contrasts were not significant, $ps \geq .028$.

In the next part, we first describe results pertaining to our calendar manipulation—its impact on accuracy and dating strategy. Then, we present results pertaining to measured variables and their relationship with dating accuracy: after describing correlations between measures, we examine memorability characteristics and then indicators of dating strategy and metacognition. Table 2 shows the proportions, counts, and dating accuracy for events across levels of each measure.

Table 2

Proportions and counts of events in levels of each measure and proportion of accurate dates within each level

Measure	Levels	Proportion	<i>N</i> _{Events}	Missing	Accurate
Dating Strategy ^{Int}	Known	.22	348	0	.80
	Reconstructed	.78	1211		.30
Event Temporal Landmark ^{Int*}	No	.92	742	0	.38
	Yes	.08	64		.84
Landmark Association ^{Exp†}	No	.89	1331	0	.35
	Direct	.11	164		.79
Frequency ^{Par}	Once	.84	1308	0	.42
	2 - 3 Times	.11	170		.37
	More Times	.05	81		.36
Regularity ^{Par}	No	.92	1429	3	.40
	Yes	.08	127		.55
Importance ^{Par}	Low	.23	356	0	.31
	Moderate	.42	661		.38
	High	.35	542		.53
Sharing ^{Par}	Not Shared	.29	450	0	.33
	Few Times	.52	809		.41
	Many Times	.19	300		.55
Emotional Intensity ^{Pro}	Low	.23	325	117	.35
	Moderate	.53	767		.44
	High	.24	350		.43
Emotional Valence ^{Pro}	Negative	.15	210	147	.40
	Ambivalent	.17	243		.37
	Positive	.68	959		.43
Confidence ^{Par}	No	.62	966	0	.22
	Yes	.38	593		.74
Testify under Oath ^{Par+}	No	.33	193	0	.50
	Yes	.67	394		.85
Self/Other ^{Exp}	Self	.91	1413	0	.43
	Other	.09	146		.23
Single Day/Extended ^{Exp}	Single Day	.85	1319	0	.37
	Extended	.15	240		.65

Note. Int = interviewer coded. Par = participant evaluated. Pro = proxy evaluated. Exp = experimenter coded. * = only data from a subset of participants in the enhanced calendar condition (*n* = 806). † = events that were also temporal landmarks were excluded (subset *n* = 1495). + = only data from a subset of participants who indicated “Yes” for initial confidence.

Calendar

We found no significant differences in dating accuracy between participants who used the enhanced calendar and participants who did not use the enhanced calendar, $OR = 1.04$ [0.70, 1.55], $z = 0.19$, $p = .852$. Participants who received the enhanced calendar for the interview self-generated temporal landmarks before the dating task. In some cases, the dated events presented at the interview were also temporal landmarks, which enabled us to assess the accuracy of dating temporal landmarks. Landmark events were dated more accurately than events that were not temporal landmarks, and the effect was large, $OR = 7.69$ [3.70, 16.00], $z = 5.46$, $p < .001$ (Table 2).

Were there differences in dating strategy between the calendar conditions? We first examined whether dated events that were temporal landmarks were events with known dates. This was the case for 92% of events that were temporal landmarks (59 out of the 64). Given that some temporal landmarks generated by participants in the calendar condition overlapped with dated events, were there more events with known dates in the calendar condition than in the no calendar condition? In the calendar condition, 24% of events were coded as having known dates compared to 20% of such events in the no calendar condition, but the difference was not significant, $\chi^2(1, N = 1559) = 4.07$, $p = .044$.

Next, was landmark association a more frequently used strategy for date reconstruction in the calendar condition? The proportions of events for which participants mentioned a direct association with a landmark was, in fact, lower in the calendar condition (9%) than in the no calendar condition (12%), but the difference was not significant, $\chi^2(1, N = 1495) = 3.25$, $p = .071$.

Correlation between Variables

The full correlation matrix for all measured variables is provided on the [OSF](#) (Table SM1). There was a weak (and rather trivial) correlation between event frequency and

regularity, indicating that events that occurred several times during the dating period were more often regular events, $r(1556) = .25, p < .001$. Event importance indicated a moderate correlation with sharing: more important events were more frequently shared, $r(1559) = .42, p < .001$. Important events were also more frequently emotionally intense, although this correlation was weak, $r(1442) = .23, p < .001$. Finally, there were noteworthy correlations between confidence and measures of dating strategy and event importance. Participants were more frequently confident when they knew (rather than reconstructed) the event date, and this correlation was moderate, $r(1559) = .46, p < .001$, when they reconstructed the date via a direct association with a temporal landmark, $r(1559) = .26, p < .001$, and when the dated events were more important, $r(1559) = .21, p < .001$, although the latter two associations were weak.

Memorability Characteristics Model

The memorability characteristics model included the following predictors: recency, single day/extended, self/other, regularity, importance, sharing, emotional intensity, emotional valence, and event frequency. The strongest predictor in this model was event embedded within an extended period (compared to single-day event) with a large effect size, $OR = 4.41 [3.01, 6.46], z = 7.60, p < .001$, followed by self-experienced (compared to other-experienced) event with a moderate effect size, $OR = 3.99 [2.30, 6.93], z = 4.91, p < .001$. Regular events were also more frequently dated accurately than unique or irregular events, and the effect was small, $OR = 2.17 [1.27, 3.72], z = 2.84, p = .004$. Event importance was associated with increased dating accuracy only for events that were rated as highly important (compared to moderately important events), and this effect was small, $OR = 1.83 [1.34, 2.51], z = 3.76, p < .001$. Finally, the odds of accurate dating estimate increased with each day, but the effect of recency was negligible, $OR = 1.02 [1.01, 1.04], z = 4.51, p < .001$ (see the right panel of Figure 1). Moderate (compared to low) event importance, any level of sharing, emotional

intensity, emotional valence, and frequency, were not significantly associated with dating accuracy, $ps \geq .037$ (complete model results are provided on the [OSF](#), Table SM2).

For regularity, we were also interested to see if temporal schemata would have an impact on day of the week estimates. Therefore, we ran a univariate analysis with regularity as predictor and day of the week estimate accuracy as the dependent variable. We found a small effect: day of the week estimates were more frequently accurate for regular events than for unique or irregular events, $OR = 2.57 [1.59, 4.17]$, $z = 3.83$, $p < .001$.

Dating Strategy and Metacognition Model

The model contained confidence, dating strategy (date known/reconstructed) and event association with a temporal landmark (note that landmark association was coded in the whole sample based on the dating protocols; for participants in the enhanced calendar condition, personal events that were also listed as landmarks were excluded from this analysis). All predictors in this model were significant with large effect sizes. The strongest predictor of dating accuracy was confidence in the final date estimate, $OR = 6.79 [5.06, 9.11]$, $z = 12.77$, $p < .001$, followed by event association with a temporal landmark, $OR = 4.76 [2.87, 7.88]$, $z = 6.06$, $p < .001$, and knowing the date (compared to reconstructing it), $OR = 4.70 [3.22, 6.85]$, $z = 8.05$, $p < .001$.

For confidence, we conducted two further univariate analyses. In the first one, we examined the association between day of the week estimates and confidence in these estimates, and we found a large effect, $OR = 9.06 [6.99, 11.75]$, $z = 16.64$, $p < .001$. Participants who indicated confidence in the day of the week were more frequently accurate in these estimates than participants who indicated that they may have confused the day of the week. Next, out of those participants who indicated confidence in their final date estimate, those who said they would be willing to testify under oath for the date were more frequently

accurate than those who said that would not be willing to testify, and the effect was large, $OR = 6.91 [4.30, 11.09]$, $z = 8.01$, $p < .001$ (Table 2).

Discussion

For recent events, we replicated the typical error performance found in previous diary studies (e.g., Thompson et al., 1996). Contrary to our expectations, we did not find a significant effect of the use of the enhanced calendar on dating accuracy—generating temporal landmarks before the dating task did not improve dating performance.

For measured indicators of accuracy, we found that dating was most accurate when participants: (1) stated that they were confident in the date; (2) used a direct relation to a temporal landmark for date reconstruction; (3) knew the date; and when the event was embedded within an extended period (this effect likely occurred because extended events more frequently served as temporal landmarks). Personal involvement indicated a moderate benefit for dating accuracy, and slightly higher dating accuracy was associated with regular and highly important events. We found no differences in dating accuracy based on event frequency, sharing, or emotional valence, and we found only a negligible effect for recency. Before we further develop our discussion of indicators of dating accuracy and their practical implications, we examined the generalizability of these indicators in Study 2 focusing on remote events.

Study 2: Remote Events

Method

The design and general method of Study 2 was the same as Study 1. The dependent variables were (1) dating accuracy (i.e., the accuracy of the month and year estimates; no/yes), and (2) accuracy of the month estimate (due to the different scale of temporal schemata for remote events; no/yes). For the sake of brevity, in the next parts, we only highlight methodological specifications and differences from Study 1.

Participants

Forty couples aged 25 to 46 years ($M = 33.2$, $SD = 4.4$; all White ethnicity) were recruited using flyers at public places and snowball sampling. Due to the reference period (2005–2008), the inclusion criterion specified that the couples needed to live together at least since 2005 (range from 1992 to 2006, $M = 2001$, $Med = 2002$; the couple who indicated the beginning of the relationship in 2006 knew each other well before that so was not excluded). The role of participants (18 self-identified as females and 22 as males) was to attend an interview about everyday memory; the role of proxies (22 self-identified as females and 18 as males; 5 completed high school and 35 had university education) was to retrieve events from their partner's life for which the proxies could verify the month and year in which they happened (see Procedure for further details).

Materials and Procedure

Retrieval of Personal Events. The task of proxies was to (secretly) retrieve a total of 28 events from their partner's life: 24 events from the years 2006 and 2007 (approximately 12 events from each year), two events from 2005, and two events from 2008. Proxies were provided with a series of instructions. (1) Focus on memorable events, events where the date (i.e., the month and year) could be verified (via e-mails, photographs, diaries, or other documents), and provide event descriptions specific enough for the participants to recognize the event as unique. Proxies provided short event descriptions and sometimes also more elaborated notes that were used during the interview in case the participants had difficulty recognizing the event. (2) Note any overlapping themes in the event descriptions. For example, "trip to Jeseníky mountains" could be a unique trip, but if "marriage proposal" occurred during this event, this should be noted in the description. (3) Omit events with obvious dates (e.g., own birthdays), and for events extending over several months, select a specific part. (4) Avoid events potentially sensitive for the partner or exclude sensitive

information into the event description. (5) Avoid using temporal hints in event descriptions and try to select events distributed across the period (i.e., avoid too many events from the same month and year).

A. N. checked that the events complied with the instructions and asked the proxies if all dates had been verified. Events for which date verification was not available or the proxy indicated low confidence in the date were excluded. The final number of events presented for an interview ranged between 16 and 28 ($M = 25.5$, $SD = 3.6$, $Med = 27$) across participants. Proxies then received written instructions and provided a series of ratings. They were asked how accurate they thought their partners (interviewed participants) would be in estimating (1) the month for each event (yes, accurate month/up to one month error/up to two months error/up to three months error/error may be more than four months) and (2) the year for each event (yes, accurate year/up to one year error/up to two years error/up to three years error/four years error). Proxies also provided estimates of (3) importance of each event for their partners (low/moderate/high). Eight public events were added to the personal events (for further details, see [OSF](#)); these results were not of central interest in this study and are not reported here (see Neusar et al., 2011; Neusar, 2012).

Calendar and No Calendar Conditions. Participants in the calendar condition received a large calendar, which provided space for a visual organization of events across months of the four years of the reference period, and additionally generated temporal landmarks before the dating task. Participants in the no calendar condition received a list of event descriptions on a sheet of paper with a blank column for the date estimates and were not asked to generate temporal landmarks before the dating task.

Interview. All interviews were conducted in 2011 between January 4 and November 20 ($M = \text{June 23}$, $Med = \text{May 30}$). There was only one calendar used in the calendar condition, which covered all months of the four-year period (2005–2008 inclusive). The interview

procedure was similar to Study 1. Interviewers coded the dating strategy (date known/reconstructed). In the calendar condition, interviewers also marked dated events that corresponded to generated temporal landmarks. In this study, participants were not asked to evaluate frequency or regularity of the dated events—the selected events were supposed to be unique, and we thought that such ratings for events that happened several years ago might not be reliable.

Event Characteristics and Confidence Ratings. Following the dating task, participants provided ratings for the degree of: (1) personal importance (low/moderate/high), (2) vividness (i.e., the amount of detail; low/moderate/high), and (3) uniqueness of each event (low/moderate/high). For confidence, participants were asked if they thought their estimate of the month was likely accurate (yes, accurate month/up to one month error/up to two months error/up to three months error/error may be more than four months), and similarly they were asked if their recall of the year was likely accurate (yes, accurate year/up to one year error/up to two years error/up to three years error/four years error). Participants who answered “yes” to the accuracy of both month and year estimates were asked if they would be willing to testify under oath for the date (no/yes).

Measures

We measured eleven variables. (1) Recency was the number of months between each event and the interview. (2) Dating strategy (date known/ reconstructed) and (3) event overlap with a temporal landmark (for participants in the calendar condition; no/yes) were coded by the interviewers.

Participants evaluated the (4) importance, (5) vividness (detail of event memory), and (6) uniqueness of each event (low/moderate/high). As for confidence, participants were first asked if they were certain that their month estimate was accurate (yes, accurate month) or whether their month estimate could have been off by up to one, two, three, or more than four

months, and they answered the same questions for their year estimate. Based on these uncertainty ratings, we created a measure of (7) confidence (no/yes): participants received a “yes” code when they were certain in both their month and year estimates and a “no” code when they indicated any uncertainty. Participants who expressed confidence in their estimates were asked if they were willing to (8) testify under oath for the accuracy of the date estimate (no/yes). The final three variables: (9) self/other-events, (10) single-day/extended events, and (11) association with a temporal landmark were coded as in Study 1.

Statistical Analyses

We computed a boundary p -value based on the 27 p -values from all tests examining dating accuracy (Benjamini & Hochberg, 1995); therefore, only p -values $\leq .004$ were considered statistically significant.

Results

Patterns of Dating Error

The left panel of Figure 2 illustrates the patterns of dating error, including the typical accumulations of errors in multiples of 12 months indicating the confusion of years. Dating errors were evenly distributed between backward telescoping (i.e., dating events as occurring in more distant past, 48%) and forward telescoping (i.e., dating events as occurring in more recent past, 52%).

Figure 2

Distribution of dating errors (left) and absolute dating error and recency (right)

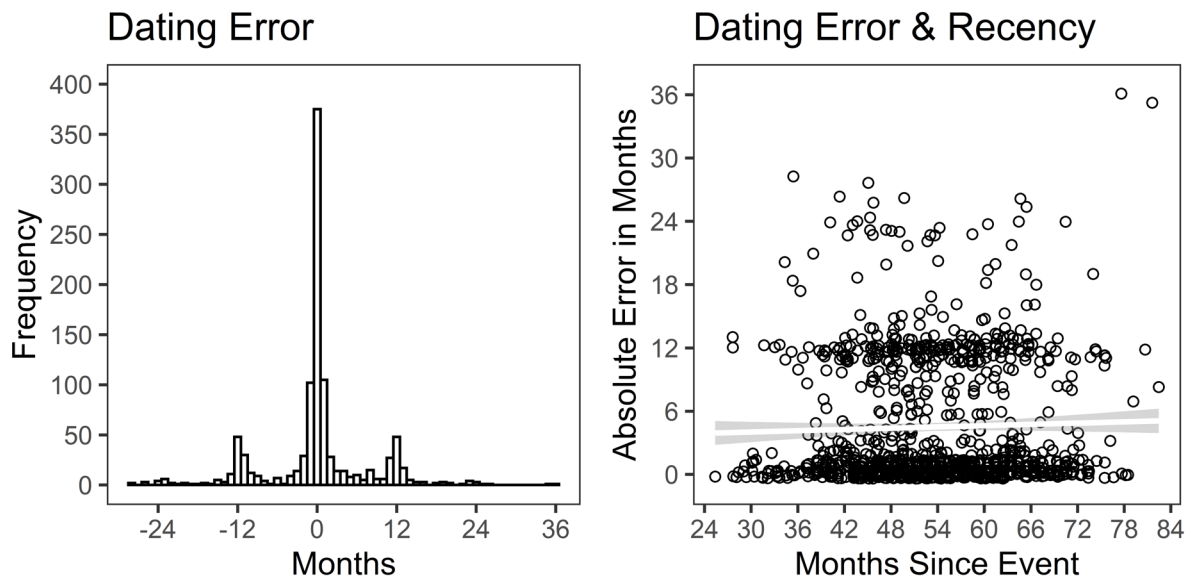


Table 3 shows proportions of accurate estimates and confusion across months in a year. The confusion matrix again illustrates how temporal schemata guide estimation: month estimates were most frequently accurate, and confusions most frequently occurred across adjacent months. There also seems to be a more fine-grained division of the year based on the school year and cultural traditions—events that occurred in the school holiday months (full months of July and August) tended to get confused within the holiday period rather than outside the period and events that occurred in December were dated more accurately than other events. We assessed the accuracy of the month estimates based on the month of the event in an analysis where we coded months with successive difference contrasts (i.e., we compared the odds of accurate month estimates between events that happened in January and February; February and March, etc.). There were two significant contrasts: the month estimates were more accurate for events that happened in December compared to November with a large effect size, $OR = 4.00 [1.87, 8.54], z = 3.58, p < .001$; and December compared to January with a medium effect size, $OR = 3.43 [1.53, 7.70], z = 2.99, p = .003$; other contrasts were not significant, $ps \geq .063$.

In the next part, we present results of dating accuracy analyses for event recency and measured indicators. Table 4 shows the proportions, counts, and dating accuracy for events across levels of each measure.

Table 3

Month of the year confusion matrix

Month	N_{Events}	Estimated Month											
		1	2	3	4	5	6	7	8	9	10	11	12
1	52	.53	.12	.09	.01	.02	.00	.01	.00	.05	.01	.03	.01
2	51	.16	.47	.08	.04	.02	.00	.00	.01	.02	.07	.02	.00
3	56	.06	.16	.37	.18	.03	.02	.00	.00	.03	.01	.00	.00
4	63	.00	.02	.14	.33	.13	.05	.04	.04	.03	.01	.00	.00
5	73	.02	.04	.08	.16	.32	.13	.01	.02	.03	.01	.05	.00
6	105	.00	.02	.03	.06	.17	.46	.14	.03	.05	.03	.02	.03
7	140	.00	.00	.02	.04	.09	.17	.54	.25	.07	.03	.00	.00
8	124	.00	.00	.02	.03	.06	.09	.20	.51	.13	.06	.03	.01
9	116	.02	.02	.05	.06	.10	.04	.04	.11	.43	.23	.08	.00
10	85	.04	.04	.03	.03	.01	.02	.02	.03	.15	.40	.19	.04
11	66	.02	.00	.08	.01	.04	.02	.00	.01	.01	.13	.48	.13
12	87	.14	.10	.03	.03	.00	.00	.00	.01	.02	.01	.10	.78

Table 4

Proportions and counts of events in levels of each measure and proportion of accurate dates within each level

Measure	Levels	Proportion	<i>N</i> _{Events}	Missing	Accurate
Dating Strategy ^{Int}	Known	.16	164	0	.70
	Reconstructed	.84	854		.31
Event Temporal Landmark ^{Int*}	No	.86	427	0	.33
	Yes	.14	67		.66
Landmark Association ^{Exp†}	No	.86	781	0	.29
	Yes	.14	129		.56
Importance ^{Par}	Low	.30	305	0	.26
	Moderate	.40	412		.35
	High	.30	301		.50
Vividness ^{Par}	Low	.19	191	0	.25
	Moderate	.30	309		.31
	High	.51	518		.45
Uniqueness ^{Par}	Low	.22	227	0	.27
	Moderate	.33	337		.31
	High	.45	454		.46
Confidence ^{Par}	No	.74	750	0	.26
	Yes	.26	268		.68
Testify under Oath ^{Par+}	No	.26	69	0	.51
	Yes	.74	197		.74
Self/Other Event ^{Exp}	Self	.86	875	0	.38
	Other	.14	143		.28
Single/ Extended Event ^{Exp}	Single Day	.72	732	0	.35
	Extended	.28	286		.42

Note. Int = interviewer coded. Par = participant evaluated. Exp = experimenter coded. * = only data from a subset of participants in the calendar condition ($n = 494$). † = events that were also temporal landmarks were excluded (subset $n = 910$). + = only data from a subset of participants who indicated “Yes” for initial confidence.

Calendar

We found no significant calendar effect on dating accuracy, $OR = 1.07 [0.67, 1.72]$, $z = 0.29$, $p = .771$. Participants in the calendar condition self-generated temporal landmarks before the dating task. We assessed the accuracy of dating temporal landmarks for events that corresponded to dated events (only for participants in the calendar condition). Landmark

events were dated more accurately than non-landmark events, and the effect was moderate, $OR = 3.75 [2.08, 6.76]$, $z = 4.40$, $p < .001$ (Table 4).

Were there differences in dating strategy between participants who used and did not use the calendar? We examined whether dated events that were temporal landmarks were events with known dates. This was the case for 74 of 108 or 69% of temporal landmarks. We then examined potential differences in dating strategy related to the generation of temporal landmarks in the calendar condition (e.g., temporal landmarks are typically events with known dates and in some cases overlapped with dated events). We found significant differences: 22% of events in the calendar condition compared to 11% in the no calendar condition were coded as having known dates, $\chi^2(1, N = 1048) = 21.08$, $p < .001$.

Was dating via a direct association with a temporal landmark a more frequent strategy of date reconstruction in the calendar condition, where participants generated temporal landmarks before the dating task? As in Study 1, we found differences in the opposite direction that were significant: 79% events in the calendar condition and 92% in the no calendar condition were dated via a direct association with a temporal landmark, $\chi^2(1, N = 910) = 28.37$, $p < .001$.

Correlations between Variables

The full correlation matrix for all measured variables can be viewed on the [OSF](#) (Table SM3). Significant correlations with dating strategy indicated that dates of more important, vivid, and unique events were more frequently known than reconstructed, although these associations were weak, $r(1018) = .22-.27$, $p \leq .001$. Dates of more vividly recalled and unique events were more frequently reconstructed via a direct association with a temporal landmark, but these associations were negligible, $r(1018) = .08-.11$, $p \leq .015$. There were also significant weak-to-moderate associations between event importance, vividness, and uniqueness, $r(1018) = .26-.60$, $p \leq .001$. Self-events and events embedded within extended

periods were recalled more vividly, $r(1018) = .12-.20, p \leq .001$, and there seemed to be a small difference in single day/extended events: self-events were more frequently embedded within extended periods, $r(1018) = .11, p = .001$, although all these associations were weak. Finally, there were again significant (though small) correlations between confidence and all measured variables except for recency and single day/extended events. Participants were more confident when they knew the date or reconstructed it via a direct connection with a temporal landmark, when the event was more important, unique, recalled vividly, and when the event was self-experienced, $r(1018) = .09-.39, p \leq .010$.

Memorability Characteristics Model

The model contained recency, uniqueness, vividness, and importance as predictors. High vividness (compared to moderate vividness) was the only significant predictor in this model, and the effect was small, $OR = 1.71 [1.19, 2.45], z = 1.85, p = .004$. Recency (the right panel of Figure 2), uniqueness, and importance were not significant, $ps \geq .049$ (complete model results are available on the [OSF](#), Table SM4).

Dating Strategy and Metacognition Model

The model contained confidence, dating strategy (date known/reconstructed) and event association with a temporal landmark (note that landmark association was coded for events in both conditions based on the dating protocols). The strongest predictor of dating accuracy was confidence in the final date estimate, $OR = 4.06 [2.87, 5.74], z = 7.93, p < .001$, followed by knowing the date (compared to reconstructing it) with a small effect size, $OR = 2.91 [1.91, 4.45], z = 4.95, p < .001$, and event association with a temporal landmark, $OR = 2.06 [1.35, 3.15], z = 3.35, p < .001$.

For confidence, we again examined the association between confidence and estimates of the month. Participants who were confident in their estimates of the month were more accurate in these estimates than participants who indicated that they may have made an error,

and the effect was large, $OR = 6.73 [4.94, 9.18]$, $z = 12.09$, $p < .001$. Finally, willingness to testify under oath in the sub-sample of participants who indicated confidence in their date estimates was associated with higher dating accuracy, and the effect was moderate, $OR = 3.59 [1.78, 7.22]$, $z = 3.58$, $p < .001$ (Table 4).

Discussion

For remote events, we again replicated the typical error performance indicating the involvement of temporal schemata on the scale of months. The strongest measured indicator of accuracy was confidence. Knowing the date (instead of reconstructing it), event association with a temporal landmark, and high vividness were also associated with higher dating accuracy, but these effects were small. As in Study 1, we found no significant effect for the use of the calendar and temporal landmark generation preceding the dating task.

General Discussion

Our examination of the retrieval of temporal estimates revealed three general moderate-to-strong indicators of accuracy: dating confidence, direct retrieval of the date, and date reconstruction via a direct association with a temporal landmark. These indicators reflect the retrieval process, its influence on confidence judgments, and the organization of autobiographical memory. Direct retrieval more likely occurs for highly memorable events and has been found to be indicative of accuracy (especially in free recall; Robinson et al., 1997). The retrieval process, in turn, influences confidence judgments, as people may be more confident in recall that came with ease (Kelley & Lindsay, 1993). Verbal protocols additionally reveal the organization of autobiographical memory, where dates of more mundane events are retrieved thanks to their links to landmark events (e.g., Brown, 2016; Shum, 1998).

The fourth strong indicator of accuracy—the actual day of the week (for recent events) or month (for remote events) the event happened—was powerful only when accuracy was

considered in calendar units of time (i.e., day of the week or month but not the actual date). That is, participants most frequently remembered that a recent event happened on Wednesday or that a remote event happened in April, although they were not as accurate in the precise dating. The second most frequent response was a confusion of adjacent units—saying that an event happened on Tuesday or Thursday when it actually happened on Wednesday or saying that an event happened in March or May when it actually happened in April (e.g., Betz & Skowronski, 1997; Gibbons & Thompson, 2001; Larsen et al., 1995).

This dating error pattern likely reflects a higher level of organization of autobiographical memory aligned with the calendar used in Western cultures that forms the basis of our knowledge about when events occur (e.g., Thompson et al., 1996). Temporal organization according to a calendar includes unit boundaries, such as the division of the week into a working week and weekend, or the division of the year into summer school holiday and school year. The dating error pattern reflects these divisions—confusions are less frequent across calendar unit boundaries. Specifically, confusions between Sunday and Monday were less frequent than confusions between Saturday and Sunday or between Monday and Tuesday; and confusions between December and January were less frequent than confusions between January and February. This organization structure (including the dating error patterns) is consistent with Estes's (1985) hierarchical model of memory and the perturbation processes that occur across attributes of remembered material. According to the model (originally developed for short-term memory tasks), the content of stimuli (words or events) is encoded along with its context—a general context such as the event detail, the overall experience, or lifetime period within a sequence of other event details, experiences, or lifetime periods, and a local context such as the position among other events including links to adjacent events and relations to boundary (anchoring) events (see also Kahana, 1996; Lohnas et al., 2015; Polyn et al., 2009).

We also assessed phenomenological characteristics of personal events as potential indicators of accuracy. High subjective event importance, direct experience, events embedded within extended periods, frequent sharing, and at least moderate emotional intensity were positively associated with dating accuracy for recent events, and high vividness was positively associated with dating accuracy of remote events. However, these associations were rather weak. It is likely that the above-mentioned phenomenological characteristics reflect memorability of other aspects of events than time (see Wagenaar, 1986).

We did not find any meaningful association between temporal recency of events and dating accuracy. It is likely that after a delay (a minimum of two weeks in Study 1), other aspects of memory for time, such as temporal schemata and event associations, rather than absolute recency, form the basis of date estimation.

Finally, we did not find that participants' dating accuracy would be affected by the use of the calendar and by retrieval of temporal landmark events before the dating task, although we did find consistent differences in dating strategy in both studies. Specifically, participants in the calendar condition more frequently "knew" dates of events and less frequently reconstructed dates via direct association with a temporal landmark. We believe that the interpretation of these differences relates to the landmark generation task: participants in the calendar condition likely *appeared* to know the dates more frequently because landmarks were already present in their calendars. Although we instructed participants to "think aloud," it is possible that participants did not mention landmark association and instead rapidly provided the date estimate. In addition, participants in the non-calendar conditions frequently spontaneously retrieved temporal landmarks during the dating task, and participants in all conditions were prompted by interviewers to think about connections between events that could be helpful for dating. It is possible that this interviewing style was sufficiently supportive for participants regardless of condition, thus diminishing the impact of the calendar

and upfront landmark generation. Calendar and visual timeline techniques likely capitalize on providing a structure to the recall task and therefore mainly facilitate the recall of specific details and their sequences within experiences (e.g., Hope et al., 2019). The recall of temporal information is already a structured process, hence the lack of the calendar benefit for dating accuracy (for a detailed review discussing mixed findings regarding calendar effects, see Glasner & Van der Vaart, 2009).

Limitations and Future Directions

We would like to consider three main limitations of this research. First, our approach to measuring confidence did not allow the examination of a fine-grained relationship between confidence and accuracy, which would require a much larger sample size. We intended to minimize demands on the interpretation of confidence rating instructions (and recent research suggests that participant interpretations of these instructions vary, e.g., Mansour, 2020), and we believe that our two-step approach (i.e., asking about confidence in the estimate and then asking whether participants would be willing to testify under oath) effectively separated highly confident participants, whose responses were accurate in 74% – 85% of cases.

Next, we amended the temporal units that were required for the final date estimate (i.e., full date in Study 1; month and year in Study 2). It is likely that events gradually lose their association with finer units of temporal schemata, but based on our data, we cannot learn about the boundaries of such associations. Therefore, further research is necessary to examine whether, for example, events that are four or eight months old retain their week schema that guides day of the week estimates, or whether the week schema association disappears.

Finally, our participants came from a relatively homogeneous group of young, white, highly educated adults from a Western culture. This presents limitations to generalizability of our findings into cultural backgrounds where people's lives may be less connected to calendar

units and more determined by ecological cycles (e.g., Aveni, 1998; Wang & Brockmeier, 2002).

Practical Implications

When looking for indicators of accuracy of temporal estimates, attributes of events that are linked to the organization of memory for personal events and metacognitive evaluation of the estimate seem to be the most reliable. The findings of the current study suggest that verbal protocols can serve as a valuable source for the evaluation of accuracy: when directly retrieved, or when retrieved via a direct association with another well-remembered event, temporal estimates tend to be accurate. Therefore, investigative interviewers should capitalize on this source and request that interviewees “think aloud” or provide descriptions of how they arrived at an estimate when retrieving temporal information. Next, we found a strong relationship between confidence and dating accuracy (see Wixted & Wells, 2017). Therefore, in line with recommendations from other fields of applied memory (e.g., Wells et al, 2020), it seems that investigators would benefit from requesting interviewees to provide confidence judgments.

Acknowledgements

The authors would like to thank Gabriela Jiskrová, Albert Kšíňan and Josef Kundrát for their help with data collection, Vojtěch Zajíček for his help with inter-rater coding, and all participants and their proxies who took part in this research for their time.

References

- Auguie, B. (2017). *gridExtra: Miscellaneous Functions for "Grid" Graphics*. R package version 2.3. <https://CRAN.R-project.org/package=gridExtra>
- Aveni, A. (1998). Time's empire. *The Wilson Quarterly*, 22, 44-57.
http://archive.wilsonquarterly.com/sites/default/files/articles/WQ_VOL22_SU_1998_Article_03_1.pdf
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67, 1-48.
<https://doi.org/10.18637/jss.v067.i01>
- Belli, R. F., & Callegaro, M. (2009). The Emergence of Calendar Interviewing. A Theoretical and Empirical Rationale. In R. F. Belli, F. P. Stafford & D. F. Alwin (Eds.), *Calendar and time diary methods in life course research* (pp. 31-54). Sage Publications, Inc.
- Ben Malek, H., Berna, F., & D'Argembeau, A. (2017). Reconstructing the times of past and future personal events. *Memory*, 25, 1402-1411.
<https://doi.org/10.1080/09658211.2017.1310251>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: Series B (Methodological)*, 57, 289-300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Betz, A. L., & Skowronski, J. J. (1997). Self-events and other-events: temporal dating and event memory. *Memory & Cognition*, 25, 701-714.
<https://doi.org/10.3758/BF03211313>
- Bradburn, N. M. (2000). Temporal representation and event dating. In A. A. Stone, J. S. Turkkan, C. A. Bachrach, J. B. Jobe, H. S. Kurtzman & V. S. Cain (Eds.), *The science of self-report: Implications for research and practice* (pp. 49-61). Lawrence Erlbaum Associates, Inc.

- Brown, N. R. (2016). Transition theory: A minimalist perspective on the organization of autobiographical memory. *Journal of Applied Research in Memory and Cognition*, 5, 128-134. <https://doi.org/10.1016/j.jarmac.2016.03.005>
- Canty, A., & Ripley, B. D. (2021). *boot: Bootstrap R (S-Plus) Functions*. R package version 1.3-26.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20, 37-46.
- Estes, W. K. (1985). Memory for temporal information. In J. A. Michon & J. L. Jackson (Eds.), *Time, mind, and behavior* (pp. 151-168). Springer.
- Finch, W. H., Bolin, J. E., & Kelley, K. (2014). *Multilevel modeling using R*. CRC Press.
- Fisher, R. P., & Geiselman, R. E. (2010). The Cognitive Interview method of conducting police interviews: Eliciting extensive information and promoting Therapeutic Jurisprudence. *International Journal of Law and Psychiatry*, 33, 321-328. <https://doi.org/10.1016/j.ijlp.2010.09.004>
- Friedman, W. J. (1993). Memory for the time of past events. *Psychological Bulletin*, 113, 44-66. <https://doi.org/10.1037/0033-2909.113.1.44>
- Gaskell, G. D., Wright, D. B., & O'Muircheartaigh, C. A. (2000). Telescoping of landmark events: Implications for survey research. [Article]. *Public Opinion Quarterly*, 64, 77-89. <https://www.jstor.org/stable/3078842>
- Gibbons, J. A., & Thompson, C. P. (2001). Using a calendar in event dating. *Applied Cognitive Psychology*, 15, 33-44. [https://doi.org/10.1002/1099-0720\(200101/02\)15:1%3C33::AID-ACP683%3E3.0.CO;2-C](https://doi.org/10.1002/1099-0720(200101/02)15:1%3C33::AID-ACP683%3E3.0.CO;2-C)
- Glasner, T. (2011). Reconstructing event histories in standardized survey research: Cognitive mechanisms and aided recall techniques [Unpublished doctoral dissertation]. Vrije Universiteit.

- Glasner, T., & Van der Vaart, W. (2009). Applications of calendar instrument in social surveys: A review. *Quality & Quantity* 43, 333-349. <https://doi.org/10.1007/s11135-007-9129-8>
- Hope, L., Gabbert, F., Kinninger, M., Kontogianni, F., Bracey, A., & Hanger, A. (2019). Who said what and when? A timeline approach to eliciting information and intelligence about conversations, plots, and plans. *Law and Human Behavior*, 43, 263-277. <https://doi.org/10.1037/lhb0000329>
- Janssen, S. M. J., Chessa, A. G., & Murre, J. M. J. (2006). Memory for time: how people date events. *Memory & Cognition*, 34, 138-147. <https://doi.org/10.3758/BF03193393>
- Kelley, C. M., & Lindsay, D. S. (1993). Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions. *Journal of memory and language*, 32, 1-24. <https://doi.org/10.1006/jmla.1993.1001>
- Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory & Cognition*, 24, 103-109. <https://doi.org/10.3758/BF03197276>
- Kuznetsova, A., Brockhoff, P. B., & Christensen; R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82, 1-26. <https://doi.org/10.18637/jss.v082.i13>
- Larsen, S. F., Thompson, C. P., & Hansen, T. (1995). Time in autobiographical memory. In D. C. Rubin (Ed.), *Remembering our past*. Cambridge University Press.
- Lee, P. J., & Brown, N. R. (2004). The role of guessing and boundaries on date estimation biases. *Psychonomic Bulletin & Review*, 11, 748-754. <https://doi.org/10.3758/BF03196630> <https://doi.org/10.3758/BF03196630>
- Ley, P. (1972). Primacy, rated importance, and the recall of medical statements. *Journal of Health and Social Behavior*, 13, 311-317. <https://doi.org/10.2307/2136768>

- Linton, M. (2000). Transformation of memory in everyday life. In U. Neisser, & I. E. Hyman (Eds.), *Memory observed: Remembering in natural contexts* (pp. 107-118). Worth Publishers.
- Literáková E., & Neusar A. (2011). Téma události [The topic of the event]. In A. Neusar (Ed.), *Kdy se to jenom stalo? (Ne)dokonalost paměti na osobní a veřejné udalosti [When did it happen? (Im)perfection of memory for personal and public events]* (pp. 98-107). MSD.
- Loftus, E. F., & Marburger, W. (1983). Since the eruption of Mt. St. Helens, has anyone beaten you up? Improving the accuracy of retrospective reports with landmark events. *Memory & Cognition*, *11*, 114-120. <https://doi.org/10.3758/BF03213465>
- Lohnas, L. J., Polyn, S. M., & Kahana, M. J. (2015). Expanding the scope of memory search: Modeling intralist and interlist effects in free recall. *Psychological Review*, *122*, 337-363. <https://doi.org/10.1037/a0039036>
- Mansour, J. K. (2020). The confidence-accuracy relationship using scale versus other methods of assessing confidence. *Journal of Applied Research in Memory and Cognition*, *9*, 215-231. <https://doi.org/10.1016/j.jarmac.2020.01.003>
- Means, B., Mingay, D. J., Nigam, A., & Zarrow, M. (1988). A cognitive approach to enhancing health survey reports of medical visits. In M. M. Gruneberg, P. E. Morris & R. N. Sykes (Eds.), *Practical aspects of memory: Current research and issues, Vol. 1: Memory in everyday life* (pp. 537-542). John Wiley & Sons.
- Neisser, U. (1988). What is ordinary memory the memory of? In U. Neisser & E. Winograd (Eds.), *Remembering reconsidered: Ecological and traditional approaches to the study of memory* (pp. 356-373). Cambridge University Press.
- Neusar, A. (2012). Memory for unique events. Predictors of dating accuracy [Doctoral dissertation, Masaryk University].

https://www.researchgate.net/publication/276172512_Memory_for_unique_events_Predictors_of_dating_accuracy

Neusar, A., Hoferková, J., & Ježek, S. (2011). Přesnost datace mediálně známých veřejných událostí [Dating Accuracy of Well-Known Public Events]. *Mediální Studia*, 5, 120-152.

https://www.researchgate.net/publication/276173478_Presnost_datace_medialne_znamych_verejnych_udalosti

Neusar, A. & Van der Vaart, W. (2018). Indicating dating accuracy by personal events ratings: Differences between self-reports and proxy-reports. In E. Maierová, L. Viktorová, J. Suchá, & M. Dolejš (Eds.), *PhD existence 2018* (pp. 69-78). Univerzita Palackého.

https://psych.upol.cz/fileadmin/userdata/FF/katedry/pch/verejnosti/sbornik_020818.pdf

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). Task context and organization in free recall. *Neuropsychologia*, 47, 2158-2163.

<https://doi.org/10.1016/j.neuropsychologia.2009.02.013>

Prober, S. M., O'Connor, M. H., & Walsh, F. J. (2011). Australian Aboriginal peoples' seasonal knowledge: a potential basis for shared understanding in environmental management. *Ecology and Society*, 16: 12.

<http://www.ecologyandsociety.org/vol16/iss2/art12/>

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>

Ritchie, T. D., Skowronski, J. J., Walker, R. W., & Wood, S. E. (2006). Comparing two perceived characteristics of autobiographical memory: Memory detail and accessibility. *Memory*, 14, 471-485. <https://doi.org/10.1080/09658210500478434>

- Roberts, J., & Horney, J. (2010). The life event calendar method in criminological research. In J. Roberts and J. Horney, *Handbook of quantitative criminology* (pp. 289-312). Springer.
- Robinson, M. D., Johnson, J. T., & Herndon, F. (1997). Reaction time and assessments of cognitive effort as predictors of eyewitness memory accuracy and confidence. *Journal of Applied Psychology, 82*, 416-425. <https://psycnet.apa.org/doi/10.1037/0021-9010.82.3.416>
- Saraiva, R. B., van Boeijen, I., Hope, L., Horselenberg, R., Sauerland, M., & van Koppen, P. J. (2020). Eyewitness metamemory predicts identification performance in biased and unbiased line-ups. *Legal and Criminological Psychology, 25*, 111-132. <https://doi.org/10.1111/lcrp.12166>
- Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language, 110*, 104038. doi: 10.1016/j.jml.2019.104038
- Schwarz, N. (2007). Retrospective and concurrent self-reports: The rationale for real-time data capture. In A. Stone, S. S. Shiffman, A. Atienza & L. Nebeling (Eds.), *The science of real-time data capture: Self-reports in health research* (pp. 11-26). Oxford University Press.
- Shum, M. S. (1998). The role of temporal landmarks in autobiographical memory processes. *Psychological Bulletin, 124*, 423-442. <https://doi.org/10.1037/0033-2909.124.3.423>
- Skowronski, J. J., Betz, A. L., Thompson, C. P., & Shannon, L. (1991). Social memory in everyday life: Recall of self-events and other-events. *Journal of Personality and Social Psychology, 60*, 831-843. <https://doi.org/10.1037/0022-3514.60.6.831>
- Skowronski, J. J., Betz, A. L., Thompson, C. P., Walker, R. W., & Shannon, L. (1994). The Impact of Differing Memory Domains on Event-Dating Processes in Self and Proxy

- Reports. In N. Schwarz & S. Sudman (Eds.), *Autobiographical memory and the validity of retrospective reports* (pp. 217-231). Springer-Verlag.
- Smith, J. P., & Thomas, D. (2003). Remembrances of things past: Test–retest reliability of retrospective migration histories. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *166*, 23-49. <https://doi.org/10.1111/1467-985X.00257>
- Sutin, A. R., & Robins, R. W. (2007). Phenomenology of autobiographical memories: The memory experiences questionnaire. *Memory*, *15*, 390-411. <https://doi.org/10.1080/09658210701256654>
- Rubin, D. C., & Kozin, M. (1984). Vivid memories. *Cognition*, *16*, 81-95. [https://doi.org/10.1016/0010-0277\(84\)90037-4](https://doi.org/10.1016/0010-0277(84)90037-4)
- Rubin, D. C., Schrauf, R. W., & Greenberg, D. L. (2003). Belief and Recollection of Autobiographical Memories. *Memory & Cognition*, *31*, 887-901. <https://doi.org/10.3758/BF03196443>
- Thompson, C. P., Skowronski, J. J., & Betz, A. L. (1993). The use of partial temporal information in dating personal events. *Memory & Cognition*, *21*, 352-360. <https://doi.org/10.3758/BF03208267>
- Thompson, C. P., Skowronski, J. S., Larsen, S. F., & Betz, A. L. (1996). *Autobiographical memory: Remembering what and remembering when*. Erlbaum.
- Van der Vaart, W., & Glasner, T. (2011). Personal landmarks as recall aids in survey interviews. *Field Methods*, *23*, 37-56. <https://doi.org/10.1177%2F1525822X10384367>
- Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S*, Fourth edition. Springer, New York. ISBN 0-387-95457-0. <http://www.stats.ox.ac.uk/pub/MASS4/>
- Wade, S. E., & Adams, R. B. (1990). Effects of importance and interest on recall of biographical text. *Journal of Literacy Research*, *22*, 331-353. <https://doi.org/10.1080%2F10862969009547717>

- Wagenaar, W. A. (1986). My memory: A study of autobiographical memory over six years. *Cognitive Psychology*, *18*, 225-252. [https://doi.org/10.1016/0010-0285\(86\)90013-7](https://doi.org/10.1016/0010-0285(86)90013-7)
- Wagenaar, W. A. (1994). Is memory self-serving. In U. Neisser & R. Fivush (Eds.), *The remembering self: Construction and accuracy in the self-narrative* (pp. 191-204). Cambridge University Press.
- Wang, Q., & Brockmeier, J. (2002). Autobiographical remembering as cultural practice: Understanding the interplay between memory, self and culture. *Culture & Psychology*, *8*, 45-64. <https://doi.org/10.1177%2F1354067X02008001618>
- Wells, G. L., Kovera, M. B., Douglass, A. B., Brewer, N., Meissner, C. A., & Wixted, J. T. (2020). Policy and procedure recommendations for the collection and preservation of eyewitness identification evidence. *Law and Human Behavior*, *44*, 3-36. <https://psycnet.apa.org/doi/10.1037/lhb0000359>
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag. ISBN 978-3-319-24277-4. <https://ggplot2.tidyverse.org>
- Williams, H. L., Conway, M. A., & Cohen, G. (2008). Autobiographical memory. In G. Cohen & M. A. Conway (Eds.), *Memory in the real world* (3rd ed., pp. 21-90). Psychology Press.
- Wixted, J. T., & Wells, G. L. (2017). The relationship between eyewitness confidence and identification accuracy: A new synthesis. *Psychological Science in the Public Interest*, *18*, 10-65. <https://doi.org/10.1177%2F1529100616686966>
- White, R. T. (2020). Autobiographical memory after 40 years. *Applied Cognitive Psychology*, *34*, 776-781. <https://doi.org/10.1002/acp.3660>
- Xu, S., & Lorber, M. F. (2014). Interrater agreement statistics with skewed data: Evaluation of alternatives to Cohen's kappa. *Journal of Consulting and Clinical Psychology*, *82*, 1219-1227. <http://dx.doi.org/10.1037/a0037489>

Yoshihama, M., Clum, K., Crampton, A., & Gillespie, B. (2002). Measuring the lifetime experience of domestic violence: Application of the life History Calendar method. *Violence and Victims, 17*, 297-317. <https://doi.org/10.1891/vivi.17.3.297.33663>