APPLIED COGNITIVE PSYCHOLOGY
Appl. Cognit. Psychol. **21**: 227–238 (2007)
Published online in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/acp.1338

Applying a Timeline as a Recall Aid in a Telephone Survey: A Record Check Study[†]

WANDER VAN DER VAART* and TINA GLASNER

VU University Amsterdam, The Netherlands

SUMMARY

This study examines the effects of a timeline tool that was employed as a visual recall aid for respondents in a standardised telephone survey. The timeline was tested in a split-ballot field experiment on the purchase behaviour of clients of opticians, the recall period being approximately 7 years. Optician database information was used as gold standard for recall accuracy. Respondents' retrospective reports about purchases of pairs of glasses were compared to the records regarding the price and the date of the most recent purchase and the number of purchases. In most cases, the timeline enhanced recall accuracy. Furthermore, it appeared to be especially helpful when the respondent's recall task was relatively difficult. The advantages and limitations of employing the timeline are discussed in relationship to the supposed underlying cognitive mechanisms.

In the recent years, a growing body of literature has focused on the use of timeline and calendar techniques in social and medical surveys (Belli, Shay, & Stafford, 2001; Caspi et al., 1996; Freedman, Thornton, Camburn, Alwin, & Young-DeMarcco, 1988; Sobell & Sobell, 2003; Sobell, Sobell, Leo, & Cancilla, 1988; Van der Vaart, 1996, 2004). These interrelated instruments form an alternative to the standard survey method in which chronologically ordered question lists are used. By providing the respondent with a graphical time frame (e.g. Figure 1), they facilitate access to long-term memory and have been shown to have beneficial effects on data quality. Although different versions of those instruments were developed relatively independently from each other in different fields of research, they share at least three important characteristics (for review see: Glasner & Van der Vaart, forthcoming):

(a) The instrument includes a graphical display of the time dimension. Usually, the reference period is divided into smaller time units, such as years, months or days. The size of those time units largely depends on the length of the reference period.
(b) The graphical display encompasses one or more themes or domains regarding which data are collected.

*Correspondence to: Wander van der Vaart, Department of Social Research Methodology, Faculty of Social Sciences, VU University Amsterdam, De Boelelaan 1081, 1081 HV Amsterdam, The Netherlands.
E-mail: w.van.der.vaart@fsw.vu.nl
[†]An earlier version of this paper was presented at the First European Association for Survey Research (EASR) Conference, Barcelona, 18–22 July 2005.

2003	**January**	**February**	**March**
interview			
1. age		48	
2. city & street	A'dam, Parkstr -----	----------	A'dam, Churchstr →
3. domestic situation	-- partner and son ---	-- partner -------	---------- →
4. jobs & courses	------ ING bank ----	----------	---------- →
5. personal landmarks		Indonesia	robbery

Figure 1. A truncated example of a filled out timeline as used in the present study

(c) The respondent is provided with temporal bounding cues such as public or idiosyncratic landmark events.

Several life course studies, in which event history calendars were used (Freedman et al., 1988; Yoshihama, Gillespie, Hammock, Belli, & Tolman, 2005), as well as a number of small-scale methodological studies on health timelines (Searles, Helzer, Rose, & Badger, 2002) report relatively high correspondence between retrospective calendar or timeline data and prior reports or collateral data. However, only a few studies exist that used a proper experimental design and directly compared these methods and a regular questionnaire. Generally, the experimental studies confirm the positive findings.

Firstly, an event history calendar was evaluated by means of a field experiment that was integrated in a longitudinal household study (Belli, Lee, Stafford, & Chou, 2004; Belli et al., 2001). The calendar was visible only to the interviewer and was used to administrate the responses. Comparing the calendar to a regular questionnaire regarding reports of events or states (e.g. moves, jobs, children aid, persons left residence, etc.) and the duration of states (work, unemployment, illness, etc.), the results generally favoured the calendar method. Secondly, a timeline tool was tested by one of the authors in a field experiment within a national, longitudinal survey on educational careers (Van der Vaart, 1996, 2004). In half of the sample the timeline was added to the standardised questionnaire and filled out by the respondents. The timeline enhanced data quality with respect to the number of educational courses followed, the starting year of the courses, and the entire set of types of courses taken. It appeared to be especially helpful for respondents who were confronted with a comparatively difficult recall task, for example those respondents, who had followed a great number of courses.

The present study aims to improve this latter timeline instrument (Van der Vaart, 2004) and explore new applications. First, we used the timeline as a visual recall aid in a telephone mode and sent the tool to the respondents prior to the interview. Second, rather than examining life history data such as educational careers, the current study focuses on the retrieval of purchase behaviour by clients of opticians, concerning: the dates, prices and

DOI: 10.1002/acp

numbers of pairs of glasses bought. Third, in contrast to most previous studies we did not compare the respondents' retrospective reports to earlier self-reports, but to record information. Opticians' sales records were used as a gold standard for recall accuracy. The study addresses two questions:

1 Does the timeline enhance the accuracy of retrospective data when it is used as a visual recall aid in telephone surveys?
2 Does the magnitude of these effects depend on the difficulty of the recall task?

RECALL ERROR AND TIMELINE INSTRUMENTS

Retrospective data quality can be compromised by several types of recall error, like omissions, biased representations of attributes and dating errors (Bradburn, Huttenlocher, & Hedges, 1994; Prohaska, Brown, & Belli, 1998; Schwarz & Sudman, 1994; Van der Vaart, 1996, 2004). Timeline and calendar techniques aim to reduce each of these types of error by combining two classes of questioning procedures that are known to enhance recall in surveys (Sudman & Bradburn, 1974): aided recall and bounding. Aided recall procedures improve the completeness of retrospective accounts, for example by providing respondents with contextual information or memory cues (Eisenhower, Mathiowetz, & Morganstein, 1991; Van der Vaart, 1996). This may help people recall the number of events as well as specific attributes of events, like the type of training course followed (Van der Vaart, 2004). Likewise, regarding the field of application in the current study, remembering the prices of products is prone to recall error (Kemp, 1999; Kemp & Willetts, 1996) and may be enhanced by timeline cues. Bounding procedures aim to enhance the accuracy of dating past events. Bounding procedures that are applied in timelines, such as relating the target event to 'landmark events' (Loftus & Marburger, 1983) or to longer autobiographical episodes, have been shown to enhance the dating of past events (Tourangeau, Rips, & Rasinski, 2000).

Belli (1998) provided a rationale for the research practice of using timelines and calendars as aided recall procedures by referring to Conway's (1996) multi-level model of autobiographical memory. According to this model, which builds on the work of Barsalou (1988), autobiographical events are embedded in a context of ongoing life experiences. The model distinguishes three highly interrelated memory structures in which autobiographical information is stored at different levels of abstraction. On the highest level, there are thematically organised lifetime periods, which consist of very long-term extended events such as working for a certain employer or living in a certain city. Themes that can be distinguished within those lifetime periods (e.g. work and relationships) are considered to be central to the self. On a lower level memories of 'general' or 'summarized' events, which took place during those lifetime periods, can be found. Those events vary in specificity (e.g. having health problems and going on holiday). Thirdly, memories of these general events are anchored in the 'phenomenological record', the memory structure in which very specific phenomenological experiences are stored. Recent studies suggest that the general events are important in autobiographical memory, because they operate as organising representations for specific memories as well as providing access to thematic knowledge (Burt, Kemp & Conway, 2003; Conway & Pleydell-Pearce, 2000).

In line with Conway's multilevel model of autobiographical memory, our timeline was designed to incorporate both high-level retrieval cues about lifetime periods as well as more specific cues (Figure 1). As Figure 1 illustrates, the timeline was divided into time

DOI: 10.1002/acp

units of years and months and encompassed five domains: age, place of residence, domestic situation, work and education and personal landmarks. We chose those particular domains for several reasons:

1 Age reports are expected to cue age related information and to make respondents aware of the lifetime period in question.
2 In terms of Conway's (1996) model of autobiographical memory, residence (city and street), domestic situation (living arrangements), work and education, respectively are thematic domains along which autobiographical episodes are organised in memory.
3 Personal landmarks provide a temporal framework for the respondents, based on which they can date the target behaviour.

HYPOTHESES ON THE TIMELINE EFFECTS

We hypothesised that our timeline—as employed in a telephone interview—would enhance recall accuracy and that its effects would be greater when respondents were confronted with a more difficult recall task. Focusing on the field of application—the recall of purchases of pairs of glasses—resulted in three hypotheses.

Hypothesis I: the timeline hypothesis

Employing the timeline in a telephone interview in addition to the standardised question list results in greater recall accuracy than using the question list only, regarding reports on:

- The price of the most recently bought pair of glasses;
- The purchase date of the most recently bought pair of glasses;
- The total number of glasses bought within the recall period.

Hypothesis II: the task difficulty hypothesis

Recall accuracy—regarding the price and date of the latest glasses and the total number of glasses—will be lower if the respondent's recall task is more difficult. This will be the case for:

- Less salient purchases (lower prices);
- Less recent purchases (more remote dates);
- More frequent purchases (greater number of glasses bought).

Hypothesis III: the interaction hypothesis

Employing the timeline (Hypothesis I) will be more effective in enhancing recall accuracy—regarding price, date and numbers—when task difficulty (Hypothesis II) is relatively high.

METHODS

The field experiment

Registered clients of opticians were interviewed about the acquisition of optical devices—in particular pairs of glasses and lenses—and related consumer behaviour. Data collection took place in May 2004, using computer assisted telephone interviewing (CATI). The

DOI: 10.1002/acp

recall period covered the full period that was available from the optician database at that moment, that is from January 1997 to March 2004. The respondents were randomly assigned to two interview conditions: a standardised interview with or without the application of the timeline.

All respondents were sent an advance letter and were promised a small incentive for participating in the study. Together with the letter, the respondents received show cards that depicted the answer alternatives of a number of questions (that were not related to the timeline). Respondents in the timeline condition were also sent a timeline accompanied by a short instruction, which included a filled-out example. The graphical design of the timeline consisted of one grid for each year (see Figure 1). The seven grids were printed one below the other on two sheets altogether. Respondents were requested to mark their main activities in the domains using brief phrases rather than detailed descriptions. The instruction only revealed that the timeline and the show cards would be used during the interview.

Respondents in both conditions were assigned randomly to the interviewers. The interviews were conducted by nine female students, between 20 and 24 years of age; all had limited interviewing experience. They were given a 2-hour interview instruction and were not informed about the expected effects of the timeline. The timeline instructions were short: the interviewers should first check whether the respondent had filled out the timeline, and if not, suggest an appointment for a call-back. During the timeline interview, the interviewers asked the respondent to mark their purchases of pairs of glasses in the timeline and check the timeline while answering the questions about these purchases.

The sample

The respondents were drawn from the client database of a branch of Dutch opticians, covering about 30% of the Dutch market. In order to ensure similar environmental conditions, we selected two middle-sized towns in the same region of the Netherlands. From the optician records of these two towns (involving 11 166 clients), a random sample of 988 clients was drawn that was representative of the total database population with respect to age and sex. Due to time restrictions during the fieldwork, 589 clients were actually contacted by an interviewer, of which 471 were eligible. The number of respondents interviewed was 233, the response rate being 49.5%. The response rate was much lower in the timeline condition (38.9%) than in the regular condition (66.9%). In spite of that, the samples in both conditions did not differ with respect to sex, age, income and education. The sample was representative for the database population in terms of age and sex: the respondents were on average 49 years of age (ranging from 18 to 70 years) and 59% of the respondents were female.

Recall accuracy and task difficulty

The opticians' records served as gold standard for the establishment of the accuracy of the respondents' retrospective reports. This applied to:

1 The price of the pair of glasses;
2 The purchase date (in months);
3 The number of pairs of glasses bought.

In order to ensure a valid match between the reported product and the product from the database, several checks were built into the research design. To begin with, the respondents

DOI: 10.1002/acp

were all registered clients of the selected branch of opticians and therefore not likely to visit others opticians. Furthermore, during the interview it was made clear to the respondents that the purchase questions only pertained to the optician where they were registered. Finally, in cases where we compared features of the glasses (i.e. its purchase date and price) we selected only the most recently bought pair of glasses in order to prevent confusion.

The respondent's task difficulty, defined as the difficulty of recalling the requested information correctly, was established by means of the purchase information in the database. Lower prices, more remote dates and higher numbers of acquisitions were assumed to indicate a more difficult recall task.

RESULTS

Method characteristics

We identified the number of call-backs, the length of the interview and the time respondents spent completing the timeline, as process variables. The number of call-backs did not differ between the control and the timeline condition (0.19 vs. 0.13, $t(211) = 1.37$, $p = 0.17$). The duration of the interviews was related to respondents' age only ($r = -0.27$, $p < 0.001$). The difference in interview duration between the control condition (21.16 minutes) and the timeline condition (23.18 minutes), was not significant (controlling for age: $F(1, 132) = 1.21$, $p = 0.27$). According to the respondents' reports, filling out the timeline took 5–30 minutes, with an average of 12.5 minutes. Within the group of 66 respondents who returned the completed timeline as requested, the number of landmarks marked in the timeline was related to the time spent on completing the timeline ($r = 0.43$, $p < 0.01$). Thanks to the completed timelines that were returned, we could observe that, generally, the timeline was filled out as intended. Most respondents had written down brief phrases in each of the domains in the timeline. They marked on average 12 landmarks in the recall period (Van der Vaart & Glasner, 2006).

Data characteristics and recall error

The database records revealed that in the timeline condition the average price of the latest pairs of glasses was lower, and the purchases dates were more recent, as compared to the control condition (Table 1). Both groups did not differ regarding the number of pairs of glasses bought during the period January 1997–March 2004. Since the disparities in price

Table 1. Differences between the respondents in the timeline and the regular condition regarding the record purchase information

	Regular	Timeline	*p*
Price (euros)	398 $N = 97$	327 $N = 82$	0.05
Date (recency in months)	37 $N = 101$	29 $N = 84$	0.01
Number	1.33 $N = 109$	1.48 $N = 86$	0.96

DOI: 10.1002/acp

Table 2. Recall error regarding the price and date of the latest pairs of glasses and the total number of glasses bought, by the regular vs. the timeline condition

	Regular	*N*	Timeline	*N*	*p*
Price					
Proportion	0.46	82	0.25	72	0.006
Date					
Net	−0.98	100	−0.88	77	0.97
Absolute	16.14		7.06		<0.001
Number					
Net	0.37	109	0.18	85	0.17
Absolute	0.59		0.53		0.63

and dates might affect recall accuracy, we incorporated these database scores as control variables when testing the hypotheses.

Table 2 illustrates the amount of recall error as established in both interviewing conditions. Regarding the price of the purchase we calculated the proportion recall error on the record price. Concerning the dates (in months) and the numbers of purchases, we measured the net and absolute differences (i.e. signed and unsigned differences) between the recalled dates and numbers and the matching record information.

The figures in Table 2 show that for each type of information recall error is smaller in the timeline than in the regular condition. In case of the number of acquired glasses, this outcome is not significant, which can probably be attributed to the fact that 91% of all the respondents had bought two pairs of glasses, at most. Such a limited variation leaves little room for the timeline to exert effects. Regarding the prices of the latest purchase, employing the timeline almost halved the amount of recall error. The timeline also enhanced recalling the year and month of the latest purchase substantially, as far as absolute recall error is concerned. However, it did not affect telescoping: the net recall errors in dates indicate that, in both conditions, forward and backward telescoping of the purchase date cancel each other out almost fully.

The timeline hypothesis and the task difficulty hypothesis

In order to test the timeline Hypothesis (I) and the task difficulty Hypothesis (II) we performed a multiple regression analyses and incorporated respondent's age and level of education as control variables (Table 3).

According to Hypothesis I, compared to the regular questioning procedure, employing the timeline would diminish the proportion of error in the reported prices and the net and absolute amounts of error in the reported dates and numbers. This hypothesis is confirmed for error in the reported prices and—in absolute terms—the reported purchase dates (Table 3). These effects thus remain when task difficulty factors, as well as age and education, are taken into account. Telescoping effects—indicated by the net error in dates—and the reports on the number of purchases, were unaffected by the interviewing condition.

Hypothesis II predicts that recall accuracy will be lower if the respondents' recall task was more difficult. Task difficulty is presumed to be higher in case of less salient purchases (lower prices), less recent purchases (more remote dates) and more frequent purchases (greater numbers).

DOI: 10.1002/acp

Table 3. Multiple regression (standardised beta's) of the task difficulty factors (record information on saliency, recency and frequency) and the interviewing condition (regular = 0, timeline = 1) on error in the recalled price ($N = 154$), date ($N = 177$) and number ($N = 194$), controlled for age and education[a]

	p-price β	e-date β	a-date β	e-number β	a-number β
Saliency (price_r)	-0.23^{***}	0.02	0.02	n.a.	n.a.
Receny (date_r)	-0.21^{**}	-0.25^{***}	-0.41^{***}	n.a.	n.a.
Frequency (number_r)	0.23^{***}	-0.03	0.07	-0.35^{***}	-0.11
Interviewing condition (timeline = 1)	-0.24^{***}	0.07	-0.21^{***}	-0.07	-0.05
Age	-0.07	-0.04	0.13^{*}	-0.09	-0.05
Education	-0.02	-0.20^{***}	-0.12^{*}	-0.03	0.009
*R*2	0.17	0.10	0.29	0.14	0.02
F	5.09	3.09	11.03	7.65	0.76
p	0.000	0.007	0.000	0.000	0.550

p, proportion error; e, net error; a, absolute error; r, record data.
*,**,***Significant at the 0.10, the 0.05 and the 0.01 level, respectively (two-tailed) n.a. = recency and saliency cannot be related to the number of purchases.
[a]The variance inflation factor (and matching tolerance estimate) ranged between 1.06 (0.85) and 1.18 (0.95) pointing out that there is no multicollinearity. Additional checks with non-parametric analyses supported the regression outcomes.

Regarding the accuracy of the recalled price, the hypothesis was confirmed for each of the task difficulty factors (Table 3). With respect to the recalled purchase dates the hypothesis was only confirmed for recency. The earlier the purchase took place, the less accurate its date was recalled, both in term of net and absolute error. Surprisingly, the price (saliency) of this purchase did not affect the correctness of the recalled date.

Next, Table 3 shows that the absolute error in the recalled number of purchases is not related to the actual (data base) frequency of purchases. The significant, but negative beta concerning net recall error indicates that recall error is smaller, if the actual purchase frequency was higher. Given the fact that 91% of the respondents bought at the most two pairs of glasses, this probably is an artefact. Respondents who bought only one pair of glasses, or no glasses at all, are most likely to overestimate the number of purchases. On the other hand, respondents with two or more purchases produce overestimates as well as underestimates, which cancel each other out.

The control variables age and education showed some (small) effects on the accuracy of the recalled dates only. Respondents with lower education made more errors in recalling the date of their last purchase. A similar but non-significant effect is found for older respondents.

The interaction between the timeline and task difficulty

Hypothesis III states that employing the timeline would be especially effective in enhancing recall accuracy when the recall task is relatively difficult. Since there were no main effects of the timeline on the accuracy of the reported number of purchases, we restricted the analyses to prices and dates (absolute error only). The adjusted hypothesis states, that the timeline will improve recall in particular for smaller purchases (low price) and less recent purchases. In analysing these two interaction effects, we dichotomised saliency (low vs. high price) and recency (remote vs. recent date) by means of the median

DOI: 10.1002/acp

Table 4. Recall error regarding the *price* of the latest pairs of glasses bought, by the interviewing condition (regular vs. the timeline) and by task difficulty in terms of purchase saliency

	Regular	*N*	Timeline	*N*
Saliency[a] (price_r)				
Low	0.72	34	0.28	37
High	0.28	48	0.21	35
	$F = 6.66$	df = 1/147	$p = 0.011$	

Interaction effect as controlled for age, education and recency.
[a]Saliency: price_r, median = € 280, Low ≤ 280 euro.

split. Using ANCOVA, we incorporated as covariates the 'other' task difficulty variable, either as recency or saliency, as well as age and education in the analysis.

Hypothesis III was confirmed for both topics. Table 4 shows a clear interaction effect concerning the prices of the most recently bought pair of glasses. In the timeline condition the reduction in recall error is 25% for high priced glasses, while the reduction is more than twice as large for lower priced glasses.

The interaction was also strong for the reported purchase dates (Table 5). For glasses that were bought within the last $2\frac{1}{2}$ years the timeline reduced recall error by about one third, that is 2 months. That effect was substantively greater for less recent purchases. Here, the timeline reduced absolute dating error by 50%. Possible floor effects in the former group may have contributed to this interaction effect.

DISCUSSION

The outcomes of this study provide a positive answer to our research question and largely support the hypotheses regarding recall of the *price* and the *date* of the purchase:

(a) In the timeline condition, as compared to the control condition, recall accuracy was generally higher and never inferior (Hypothesis I); unexpectedly, there was no effect on telescoping (net error in dates);
(b) More difficult recall tasks, that is involving less salient, less recent (and to some extent more frequent) purchases—coincided with greater recall error (Hypothesis II);
(c) Interaction effects between the interviewing condition and task difficulty were established: the higher levels of recall accuracy in the timeline condition were especially pronounced if the recall task was relatively difficult (Hypothesis III).

Table 5. Recall error regarding the *date* of the latest pair of glasses bought, by the interviewing condition (regular vs. the timeline) and by task difficulty in terms of purchase recency

	Regular	*N*	Timeline	*N*
Receny[a] (date_r)				
Remote	23.79	56	10.67	33
Recent	6.41	44	4.36	44
	$F = 8.79$	df = 1/166	$p = 0.003$	

Interaction effect as controlled for age, education and saliency.
[a]Recency: date_r = October 2001, Remote ≥ 30 months.

DOI: 10.1002/acp

The lack of variation in the *number* of purchases limited our chances of finding significant effects on this measure.

Assets and drawbacks

The results of the current study are generally consistent with earlier findings on timeline and calendar instruments. Moreover, while our prior study (Van der Vaart, 2004) was limited to major (educational) lifetime episodes and relatively young respondents (18–30 years), here similar effects were obtained for purchase events and for a great range of ages (18–70 years). The effects of the timeline were much greater in the current study, probably because reports on pairs of glasses are characterised by lower levels of recall accuracy to begin with.

The most important difference between the current study and earlier applications was that respondents completed the timeline before the interview and had complete control over the way in which they used it as a recall aid. This had advantages as well as drawbacks. Regarding the duration of the interviews the timeline procedure appeared to be rather efficient. Having respondents fill in the timeline beforehand clearly saved time during the interview. On the other hand, in the timeline condition the response rate was substantially lower than in the regular condition. Although we could not find any evidence of biased respondent selection, it is possible that the participants in the timeline condition were more motivated than those in the control condition, which might have resulted in more accurate retrospective reports. This motivational effect has been found for aided recall procedures is general (Sudman & Bradburn, 1983). Yet, motivation alone cannot explain why the effect of the timeline was particularly great for respondents who were faced with a difficult recall task. We therefore suggest that there might be an additional effect of motivation on recall accuracy, which is complementary to the cognitive effects of the recall aid. If sending the timeline tool beforehand actually does deter respondents, one solution could be to employ the timeline in mix-mode studies that include a panel of respondents with home computers. In that case the timeline could be presented on the respondent's computer screen during telephone interviews. In addition, if the timeline produces higher data quality, then population estimates could be achieved with a smaller sample, in which case some reduction in response rate might be an acceptable trade off.

In the present study, the timeline was purposely devised as a general temporal framework that would be applicable to almost any topic. However, the effectiveness of the timeline might be improved by making it more domain-specific. The themes and events in the calendar could be specified and directly related—thematically and causally—to the target information. Research into narrative-like structures of memory, and event clusters suggest that this might enhance retrieval (Brown, 2005). In our case timeline domains could be focused on optician related issues, like associated health situations, or consequences of having bad sight, etc.

This study provided further evidence that timeline methods may enhance recall accuracy in surveys. The outcomes indicate that timeline applications are not limited to a small range of issues or specific research populations. Although it is assumed that the effects of timeline methods may be attributed to cognitive mechanisms, more research is needed to enlarge our knowledge about the processes involved. Given the potential benefits that timeline methods can contribute to retrospective survey designs, pursuing that type of research is worth the effort.

DOI: 10.1002/acp

ACKNOWLEDGEMENTS

The authors would like to thank Marjolein Nijhof for managing the fieldwork and Loyalty Lab Amsterdam for cooperating and providing the database information. This work was supported by a grant from the Faculty of Social Science, VU University Amsterdam.

REFERENCES

Barsalou, L. W. (1988). The content and organization of autobiographical memories. In U. Neisser, & E. Winograd (Eds.), *Remembering reconsidered: Ecological and traditional approaches to the study of memory* (pp. 193–243). New York: Cambridge University Press.

Belli, R. F. (1998). The structure of autobiographical memory and the event history calendar: Potential improvements in the quality of retrospective reports in surveys. *Memory, 6*, 383–406.

Belli, R. F., Lee, E. H., Stafford, F. P., & Chou, C.-H. (2004). Calendar and survey question-list methods: Association between interviewer behaviors and data quality. *Journal of Official Statistics, 20*, 185–218.

Belli, R. F., Shay, W., & Stafford, F. (2001). Event history calendars and question list surveys: A direct comparison of interviewing methods. *Public Opinion Quarterly, 65*, 45–74.

Bradburn, N., Huttenlocher, J., & Hedges, L. (1994). Telescoping and temporal memory. In N. Schwarz, & S. Sudman (Eds.), *Autobiographical memory and the validity of retrospective reports* (pp. 203–215). New York: Springer-Verlag.

Brown, N. R. (2005). On the prevalence of event clusters in autobiographical memory. *Social Cognition, 23*, 35–69.

Burt, C. D. B., Kemp, S., & Conway, M. A. (2003). Themes, events, and episodes in autobiographical memory. *Memory & Cognition, 32*, 317–325.

Caspi, A., Moffitt, T. E., Thornton, A., Freedman, D., Amell, J. W., Harrington, H., Smeijers, J., & Silva, P. A. (1996). The life history calendar: A research and clinical assessment method for collecting retrospective event-history data. *International Journal of Methods in Psychiatric Research, 6*, 101–114.

Conway, M. A. (1996). Autobiographical knowledge and autobiographical memories. In D. Rubin (Ed.), *Remembering our past: Studies in autobiographical memory* (pp. 67–93). Cambridge: Cambridge University Press.

Conway, M. A., & Pleydell-Pearce, C. W. (2000). The construction of autobiographical memories in the self-memory system. *Psychological Review, 107*, 261–288.

Eisenhower, D., Mathiowetz, N. A., & Morganstein, D. (1991). Recall error: Sources and bias reduction techniques. In P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, & S. Sudman (Eds.), *Measurement errors in surveys* (pp. 127–144). New York: John Wiley.

Freedman, D., Thornton, A., Camburn, D., Alwin, D., & Young-DeMarcco, L. (1988). The life history calendar: A technique for collecting retrospective data. *Sociological Methodology, 18*, 37–68.

Glasner, T. J., & Van der Vaart, W. (forthcoming). Applications of calendar instruments in social surveys: A review. Manuscript accepted by Quality & Quantity.

Kemp, S. (1999). An associative theory of estimating past dates and past prices. *Psychonomic Bulletin & Review, 6*, 41–56.

Kemp, S., & Willetts, K. (1996). Remembering the price of wool. *Journal of Economic Psychology, 17*, 115–125.

Loftus, E. F., & Marburger, W. (1983). Since the eruption of Mt. St. Helens, did anyone beat you up? Improving the accuracy of retrospective reports with landmark events. *Memory & Cognition, 2*, 114–120.

Prohaska, V., Brown, N., & Belli, R. F. (1998). Forward telescoping: The question matters. *Memory, 66*, 455–465.

Schwarz, N., & Sudman, S. (Eds.). (1994). *Autobiographical memory and the validity of retrospective reports*. New York: Springer-Verlag.

DOI: 10.1002/acp

Searles, J. S., Helzer, J. E., Rose, G. L., & Badger, G. J. (2002). Concurrent and retrospective reports of alcohol consumption across 30, 90 and 366 days: Interactive voice response compared with the timeline follow back. *Journal of Studies on Alcohol, 63*, 352–362.

Sobell, L. C., & Sobell, M. B. (2003). Alcohol consumption measures. In J. P. Allen, & V. B. Wilson (Eds.), *Assessing alcohol problems: A guide for clinicians and researchers* (2nd ed., pp. 75–99). Bethesda, MD: National Institute on Alcohol Abuse and Alcoholism.

Sobell, L. C., Sobell, M. B., Leo, G. I., & Cancilla, A. (1988). Reliability of a timeline method: Assessing normal drinkers' reports of recent drinking and a comparative evaluation across several populations. *British Journal of Addiction, 83*, 393–402.

Sudman, S., & Bradburn, N. M. (1974). *Response effects in surveys: A review and synthesis*. Chicago: Aldine.

Sudman, S., & Bradburn, N. M. (1983). *Asking questions: A practical guide to questionnaire design*. London: Jossey-Bass.

Tourangeau, R., Rips, L., & Rasinski, K. A. (2000). *The psychology of survey response*. Cambridge: Cambridge University Press.

Van der Vaart, W. (1996). Inquiring into the past. Data quality of responses to retrospective questions. PhD-thesis. Amsterdam: Vrije Universiteit.

Van der Vaart, W. (2004). The time-line as a device to enhance recall in standardized research interviews: A split ballot study. *Journal of Official Statistics, 20*, 301–317.

Van der Vaart, W., & Glasner, T. J. (2006). Using personal landmarks to enhance recall accuracy in survey interviews. Paper presented at the XVIth World Congress of Sociology, Durban, July 23–29, 2006.

Yoshihama, M., Gillespie, B., Hammock, A., Belli, R. F., & Tolman, R. (2005). Does the life-history calendar method facilitate the recall of domestic violence victimization? Comparison of two methods of data collection. *Social Work Research, 29*, 151–163.

APPENDIX: TARGET QUESTIONS

The target questions in this timeline field experiment were phrased as follows (identical in both conditions):

A How often did you buy a pair of glasses from your optician during the period from January 1997 to March 2004?

B When did you buy the last (or this) pair of glasses from your optician: in which year and month?

C What was the price of this last pair of glasses?

DOI: 10.1002/acp