Running Head: MEMORY AND RETROSPECTIVE EVALUATIONS

Retrospective Evaluations of Sequences:

Testing the Predictions of a Memory-based Analysis

Silvio Aldrovandi

Marie Poirier

Petko Kusev

Peter Ayton

Author Note

Silvio Aldrovandi and Marie Poirier, Department of Psychology, City University London; Petko Kusev, Department of Psychology, Kingston University London; Peter Ayton, Department of Psychology, City University London.

Silvio Aldrovandi is now at the Division of Psychology, Birmingham City University.

This research was partially supported by the ESRC grant RES-062-23-2462.

The authors would like to thank Daniel Heussen, Stephan Lewandowsky and James Nairne for their helpful comments on a previous version of the manuscript.

Experiment 3 has been previously published as a Cognitive Science Conference Proceedings paper (Aldrovandi et al., 2011).

Correspondence concerning this article should be addressed to Silvio Aldrovandi, Division of Psychology, Birmingham City University, Franchise Street, Birmingham, B42 2SU.

Email: Silvio.Aldrovandi@bcu.ac.uk

Abstract

Retrospective evaluation (RE) of event sequences are known to be biased in various ways. The present paper presents a series of studies that examined the suggestion that the moments that are the most accessible in memory at the point of RE contribute to these biases. As predicted by this memory-based analysis, Experiment 1 showed that pleasantness ratings of word lists were biased by the presentation position of a negative item and by how easy the negative information was to retrieve. Experiment 2 ruled out the hypothesis that these findings were due to the dual nature of the task called upon. Experiment 3 further manipulated the memorability of the negative items—and corresponding changes in RE were as predicted. Finally, Experiment 4 extended the findings to more complex stimuli involving event narratives. Overall, the results suggest that assessments were adjusted based on the retrieval of the most readily available information.

Keywords: retrospective evaluations, memory, on-line judgement, accessibility, order effects

Retrospective Evaluations of Sequences: Testing the Predictions of a Memory-based Analysis

It is a common experience to reflect on past events, evaluating them in hindsight. Events typically hold features we encounter in a sequential manner, as information unfolds over time; such experiences are usually called temporal sequences (Lowenstein & Prelec, 1993). Each piece of information acquired might have varied in quality, intensity or valence; nevertheless, retrospective evaluations (RE) typically involve a cohesive unitary response (e.g., Baumgartner, Sujan, & Padgett, 1997).

RE have been investigated in a number of disparate domains (e.g., Fredrickson, 2000), including for example, the evaluation of pain (Redelmeier & Kahneman, 1996) and payment sequences (Langer, Sarin, & Weber, 2005). One of the important conclusions of this work is that RE are biased. For instance, the 'Peak-End' rule proposes that the final moments and the most intense parts of an experience disproportionately affect RE (Fredrickson & Kahneman, 1993). Personality impression formation is also biased; when participants read a set of adjectives about a hypothetical character, the impression formed is usually more positive when favourable adjectives precede more unfavourable ones, rather than the reverse (Asch, 1946).

A common feature of the above findings is that certain moments or characteristics of an event weigh more heavily in RE than others (e.g., Ariely & Carmon, 2000). One overarching interpretation of these findings involves a memory-based approach; this approach suggests that the aforementioned biases are attributable to the accessibility of the biasing moments in memory (e.g., Montgomery & Unnava, 2009). Indeed, a number of authors have recently argued that a more thorough consideration of specific memory processes would benefit our understanding of a number of issues in the judgment and decision making literature (e.g., Platzer & Bröder, 2012; Weber & Johnson, 2009).

With respect to RE, consider for example the peak-end effects mentioned above; these could be understood by assuming that the peak is distinctive and hence more memorable while the end benefits from the mnemonic advantages related to the recency effect, where the last encountered information in a series is better remembered. If it is assumed that the more accessible moments at the point of RE weigh more heavily in the assessment, then biases can be predicted based on the memorability of sequence elements.

Recent findings provide support for a memory-based approach to RE. For example, Montgomery and Unnava (2009; Study 1) presented participants with a description of a vacation that included incidents of different quality, ranging from very negative (a bad sunburn) to very positive (discovering and enjoying a festival). The position of negative and positive incidents was manipulated between-subjects, so that participants either received an improving sequence or a deteriorating one. Participants were then asked to rate the vacation as a whole and to recall events. Crucially, the authors manipulated the delay between the presentation of the sequences and their evaluation and recall. In line with well-established findings in the memory literature (e.g., Bjork & Whitten, 1974), the authors found that the delay led to a decrease in recency effects. More importantly, the recall results showed that the delay lowered the accessibility of the information appearing later in the sequence; as a result, deteriorating sequences were preferred to improving ones—the opposite pattern was observed in the immediate condition.

Although these findings provide compelling evidence for a memory-based approach, an analysis reported in an influential paper by Hastie and Park (1986) may limit the generalisability and impact of this type of evidence. Hastie and Park (1986) examined the role of memory in RE and concluded that memory-based assessments are unlikely when participants are aware of the upcoming judgment task—as people are likely to be constructing their assessment as events unfold (*on-line* evaluations). Indeed, a number of studies in social cognition have found memory and judgment to be unrelated (e.g., Anderson, 1989; Lichtenstein & Srull, 1987). Less surprisingly, the authors found that when people are instead unaware of the upcoming judgment task, they must rely on the mnemonic record of the actual processed information in order to produce RE: This is because the on-line evaluation did not take place.

Acknowledging these considerations, Montgomery and Unnava (2009) suggested that situations that involve on-line evaluations may reduce the predictive value of a memory-based analysis. Indeed, in their Study 1, participants only evaluated *one* sequence of items and they were not aware of the specifics of the judgment task to follow. Hence, it is not clear if a memory-based approach could predict RE when participants are aware that an overall assessment will be required—as this situation is most likely to foster on-line evaluations.

The general aim of the present paper is to provide a stricter test of the predictions of a memory-based analysis of RE (e.g., Montgomery & Unnava, 2009; Weber & Johnson, 2009); more specifically, we investigate whether a memory-based analysis can predict RE even in situations where on-line judgment formation is not inhibited.

Also, the role that memory may play in biasing RE is under further scrutiny in a number of ways. First, for the first time, we tested whether any association between memory and RE would be maintained when participants could not anticipate a memory test. For most studies investigating the relationships between memory and judgment (e.g., Hastie & Park, 1986; Montgomery & Unnava, 2009), the dual nature of the task might have artificially induced participants to rely on the memory representation when providing a RE, when otherwise they would not do so.

Second, we used a series of memory manipulations—implemented either during or after stimulus encoding—and put forward specific predictions based on the expected memory-based contribution to RE.

Third, to our knowledge, previous studies have not controlled for variables which are known to affect memory performance (e.g., stimulus frequency and familiarity); the studies reported below included these controls. Such factors could artificially inflate or diminish the associations between memory and RE; for instance, a positive attribute might have an impact on RE because of its valence, but it may not be recalled because other items are higher in familiarity or frequency (and will hence be easier to recall).

Lastly, we present more detailed analyses of the associations between memory and RE. Previous research has generally involved simple correlations between RE ratings and global or general memory measures where each item recalled is assumed to have the same influence on judgment (e.g., Hastie & Park, 1986). To our knowledge, this paper is the first to provide an analysis of the association between recall output position and RE. Since items could be recalled in any order, we made the simplifying assumption that items recalled first were on average more readily accessible in memory. This analysis allowed us to look in further detail at the memory processes and representations that inform judgment and decision-making (e.g., Weber & Johnson, 2009) in relation to accessibility. The overweighing of specific information in RE might not be due simply to it being present (vs. absent) in memory at the time of judgment: If participants are accessing the degraded trace of the just-seen information, then a single item might be 'weighed' in judgment inversely to its degradation.

Experimental Strategy and Predictions

Participants rated the pleasantness of word lists (Experiments 1, 2 and 3) or stories told through slideshows and audio recordings (Experiment 4). After providing this judgment, a recall task was performed. Memory always followed judgment; this strategy was adopted because recall followed by judgement could artificially boost the relationship between the remembered material and RE (e.g., Schwarz & Vaughn, 2002).

Experimental sequences contained a negative item (or event), inserted within neutral material. Across experiments, the memorability of this negative item was varied in a number of ways. We also examined how judgment varies as a function of the recall of the negative information. If the negative item is recalled after RE, it seems likely that it was also available at the time of judgment. Conversely, if the negative item is not available for recall, the probability that it was available at the time of judgment is reduced. Hence, on average, ratings should be lower when the negative item is available for the memory component of the task. We further examined the relationship between RE and memory further by analysing output position in the recall task; it was hypothesised that the earlier a negative item was recalled, the stronger its impact on RE.

General Methodology

Materials

For the word list experiments, item valence was manipulated. The words (see Appendix A) were selected from the Affective Norms of English Words database (ANEW; Bradley & Lang, 1999). Negative words were low in valence (less than 3, on a 1-9 scale) and high in arousal (over 6, on a 1-9 scale); neutral item valence was middling (4.5 to 6.9) and low on the arousal scale (less than 5). The serial position in which a negative item appeared was manipulated. For instance, ‘P1’ lists included a negative item in the first position, followed by neutral items. ‘Control’ lists contained only neutral items.

Negative lists—i.e. those including a negative item—were matched on valence and arousal ratings. Control lists were matched on both valence and arousal with the neutral words included in the negative lists. Negative items were rotated, so that each one appeared in each position an equal number of times across participants; items were not repeated across trials.

For Experiment 4, a female voice presented a narrative that accompanied a series of six sequentially presented photographs/slides. The pictures were selected from the International Affective Pictures System (IAPS; Lang, Bradley, & Cuthbert, 1999). Negative slides were low in valence (less than 3, on a 1-9 scale), while neutral slides were middling (from 3.8 to 5.9; d = 4.1). The negative slides (M = 4.0) were also higher in arousal ratings than neutral slides (M = 2.4, d = 2.6).

Within-list matching between the negative (if any) and the neutral items ensured that negative and neutral words were equated on familiarity ratings (Coltheart, 1981), number of phonemes, and ‘K-F’ frequency (Kucera & Francis, 1967). For the slides, matching was based on distinctiveness and memorability ratings (Lang et al., 1999)—and the number of words and phonemes used in the narratives accompanying each slide. Every effort was made to equate the recordings of the narratives across stories. The same female voice was used, and the tone of the voice used was regular and did not express emotion. The grammatical structure of all the sentences was comparable – they usually involved only a subject and a predicate.

Procedure

Participants in Experiments 1 and 3 were individually tested via a computer program written in Authorware 7.0 (Adobe / Macromedia). Experiments 2 and 4 were run online.

In all experiments, introductory screens presented the procedure and gathered demographic data. Participants were told that the aim of the experiment was to collect normative data regarding the pleasantness of word lists (or stories). For words lists, each word appeared for one second with an inter-stimulus interval of 0.75 seconds. A series of asterisks appeared for three seconds to signal the end of list presentation. For the narratives, each slide was on screen for 4 seconds. As the recording that accompanied each slide did not exceed 3.1 seconds, brief pauses surrounded each sentence.

Participants were to attend to each series and provide an overall pleasantness rating. The ratings were on a 0-100 scale (0 = *very unpleasant*, 100 = *very pleasant*) and use of the whole range was encouraged. Participants clicked on a slider bar (with extremes of 0 and 100) on a position they felt was closest to their impression of the list/story. To limit anchoring effects (e.g., Chapman & Johnson, 2002) a sliding marker appeared on the bar, with its numerical value below, only after the first click. Participants could then adjust this initial rating; they confirmed their final rating by clicking on a "Continue" button.

The memory task was an easy version of a free recall task. After each series, participants had to type in the two elements that most readily came to mind –words or a brief description of story moments; they then pressed a continue button. Then, they could recall any further item they remembered, if any. It was stressed that it was equally fine to proceed directly to the next trial, without adding any further items, and that both the judgment and recall tasks were important. Participants were specifically asked not to neglect the rating task in order to proceed more quickly to the recall task. This modified version of the free recall task was used to reduce cognitive demands. Previous research showed that typical recall tasks, which require participants to perform an exhaustive search in memory, can mask associations between memory and judgment (e.g., Kitayama & Burnstein, 1989). Two practice trials were provided. List/story presentation order was randomised independently for each participant and no time limits were set for the rating and recall tasks.

Data Analysis

When analysing the relationship between memory and RE, in order to reduce the extent of inter-individual differences in the use of the rating scale, judgment scores were transformed. For each participant, the average pleasantness rating for the Control lists was subtracted from the pleasantness ratings for each list containing a negative item. The corrected judgment scores (J') represented how much more unpleasant each list was in comparison to the average Control list. Average J' scores were obtained for each participant, according to the negative item presentation position and whether the negative item presented in the list was recalled or not.[1]

Due to the dual nature of the task, a precautionary measure was taken in order to exclude from the analyses any participant who neglected the judgment task. Participants whose judgment scores were characterised by a standard deviation smaller than five (i.e., 5% of the 0-100 scale) were eliminated from the analyses (in total, from 0.8% to 3.1% across experiments). Moreover, for the online experiments, participants were also excluded if the time to complete the task was three times longer (or more) than the average duration of the experiment (in total, from 1.3% to 4.4% across experiments).

Experiment 1

In this experiment, the position of a negative item within the to-be-assessed list was manipulated. The predictions of a memory-based approach are straightforward. A negative item appearing in the first (primacy) or last (recency) position will tend to be more readily available in memory following list presentation. It follows that RE for such lists is predicted to be lower than RE for lists where the negative item appears in the middle positions.

Method

Participants. A total of 38 participants (21 females) aged from 19 to 55 (*M* = 40.9, *SD* = 11.4) took part in the study; they were awarded £7 for their participation.

Design, Materials and Procedure. A series of 32 six-word lists were created (see Appendix); there were eight P1, P6 and Control lists as well as four P3 and four P4 lists (i.e. eight lists with a negative item in middle positions). After each list, participants performed the (a) RE and (b) memory tasks as described in the general method.

Results and Discussion

Memory. Figure 1 (left panel) presents *mean correct recall* as a function of word position and valence. Negative items presented first or last were better recalled than negative items presented in the 3rd or 4th positions.

A 2 (valence: negative vs. neutral) × 3 (position: 1st, 3rd/4th, and 6th) repeated measures ANOVA revealed a significant main effect of valence; negative items were better recalled than neutral ones and there was no interaction between position and valence. Note that all the ANOVA-based inferential statistics for all the experiments are presented in Table 1. A main effect of position showed that P1 recall was higher than P3/4 (d = 0.94) and also that P6 recall was higher than P3/4 (d = 0.31); thus, primacy and recency effects were observed for the recall of the negative items.

RE. The *mean pleasantness rating* for the Control lists was higher—more positive—than the ratings for the other list types. Moreover, there is evidence of both primacy and recency in the ratings (see Figure 1, right panel): P1 and P6 lists were rated less positively than the lists with a negative item in the middle positions.

A one-way ANOVA with list type (P1, P3/4, P6 and Control) as the within-subject factor confirmed the above observations. Judgments for P6 or recency lists were lower than those for P3/4 lists (d = 0.22); ratings for P1 or primacy lists were lower than ratings for P3/4 lists (d = 0.31).

We also ran a subsidiary study where no memory task was used, as this allowed us to determine whether demand effects altered the retrospective judgment pattern. A total of 21 participants (11 males) volunteered to take part in this study; their ages ranged from 21 to 40 years (M = 27.9, SD = 4.9). The results showed that the RE pattern was the same when no memory task was included: Both primacy (d = 0.54) and recency effects (d = 0.65) were observed for pleasantness ratings.

Memory-RE Associations. Overall, when the negative item was recalled in the memory task, pleasantness ratings were lower (*M* = -15.7, *SD* = 10.7) than when the negative item was not recalled (*M* = -8.5, *SD* = 9.2); this relationship between memory and judgment seems evident across list types (see Table 2).

A significant main effect of memory (item recalled vs. not recalled) confirmed the above observations; moreover, the relationship between memory and judgment was observed across positions, as the list type by memory interaction was not significant.

Finally, RE results were analysed depending on the output position of the negative item in recall. Table 3 shows that the pleasantness ratings were lowest when the participants recalled the negative item as either the first or second response. A significant main effect of recall position on pleasantness ratings confirmed these observations and showed that judgments were lower when the negative item was recalled amongst the first two responses than when it was recalled amongst the last four responses (*d* = 0.77); moreover, the latter judgments were lower than those observed for lists where the negative item was not recalled at all (*d* = 0.42).

The results of Experiment 1 confirmed that a negative item presented at the beginning or the end of the series exerted a larger impact on pleasantness evaluations. Primacy and recency effects were observed also for the recall of the negative items.

Associations between memory and RE scores were observed: The pleasantness ratings for those lists where the negative item was not recalled were significantly higher (of 8.5 units on average, on a 0-100 scale) than for those lists where the negative item was recalled. Moreover, accessibility in memory of a negative item appeared to impact RE, since the lowest ratings were provided when the negative item was recalled as an early response.

Two arguments might undermine these observations. First, a third factor could be mediating the serial position effects on memory and the mirrored effect on RE, issue that we address in Experiment 3. Second, the dual nature of the task might have artificially inflated the relationships between memory and RE. When performing the evaluation task, participants knew that they had to perform a memory task subsequently; hence, participants might have rehearsed the degraded representation of the list at the time of judgment—when otherwise they would not do so (e.g., Anderson, 1989). In other words, a general alternative explanation for our findings is that having to do both judgment and memory tasks for the same lists artificially inflated the associations between memory and judgment in Experiment 1. Experiment 2 addressed this issue.

Experiment 2

This experiment called upon a *surprise* memory task. As participants could not anticipate a memory test, there is no reason for the memory component of the procedure to have an impact on the association between what is recalled and the pleasantness ratings.

Method

Participants. A total of 113 participants (66 males) took part in an internet-based experiment. Participants' age ranged from 20 to 63 years (M = 41.6, SD = 12.7); they were granted Maximiles® points in exchange for their participation.

Design, Materials and Procedure. The stimuli were 20 seven-word lists; there were five lists of each of the following types: P1, P4, P7 and Control. Each participant was presented with four trials in total; in the first trial they were always presented with a Control list randomly selected from the set of 5. The following trials contained a negative item with position rotated evenly across participants. The fourth and last trial was followed by a surprise memory test.

Participants were told that they were to assess the pleasantness of word lists. Once they had rated the 4th and last list, a screen instructed them to recall the two words that came to mind most readily from the just-presented list; on a further screen participants had a chance to enter any other word they remembered from the list.

Results and Discussion

First, we examined the difference in ratings for the trials where the negative item was recalled relative to the trials where it was not. An independent samples *t*-test with memory (negative item recalled vs. not recalled) as the between-subjects factor, $t(105) = 2.6$, $p = .011$, $d = 0.52$, confirmed that pleasantness ratings were higher for lists where the negative item was not recalled ($M = 0.3$, $SD = 20.7$)[2] relative to lists where the negative item was recalled ($M = -10.3$, $SD = 19.9$).

Next, we investigated the relationship between accessibility of the negative item and judgment (Table 3). There was a significant main effect of negative item recall position (not recalled; recalled as 1^{st} response; recalled as 2^{nd} to 4^{th} response) on pleasantness ratings; when the negative item was recalled early pleasantness ratings were lower than when it was either not recalled or recalled as a later response (both *d*s > .51). RE when the negative item was recalled as 2^{nd} to 4^{th} response did not differ from when it was not recalled at all ($p = .47$).

Overall, the correspondence between memory and judgment was clear: RE was lower when the negative item was recalled and it was lowest when the negative item was most easily accessible in memory. As this pattern was observed even if participants did not expect a memory task, it is reasonable to conclude that the memory paradigm called upon in Experiment 1 cannot be the source of memory-related effects on RE.

Experiment 3

Here, we implement a further test of the memory-based analysis by using another manipulation of the memorability of the negative item; indeed, experimental evidence which highlights correlations between memory and RE does not preclude that such correlations are attributable to other factors (e.g., vividness; Shedler & Manis, 1986; see also Moser, 1992).

Crucially, in this experiment the memorability of information was manipulated *after* its presentation, as opposed to *during* encoding. For some lists a filled delay was inserted after the item presentation. The effects of a filled delay are well established: They lead to a *reduced* recency effect. If a negative item's accessibility in memory is lowered—by presenting it at the end of a list where a filled delay follows—then its impact on the summary assessment of the list as a whole should be smaller than if its accessibility is not hindered (no delay). As accessibility of the items in memory was manipulated after stimulus encoding, no *specific, position-dependent*, delay effect would be expected for judgments if RE is based on evaluative processes that take place during list presentation.

Method

Participants. A total of 79 participants (49 males) took part in the internet-based experiment. Participants' age ranged from 28 to 65 years (*M* = 46.9, *SD* = 10.2) and they were granted Maximiles® points.

Design, Materials and Procedure. The procedure was the same as in Experiments 1 and 2, except for the delay manipulation. For half of the trials, after list presentation, participants were prompted to provide their pleasantness rating (immediate condition). For the remaining half, participants had to engage in a 10-second distractor task (delay condition). During this task, a letter was presented on the screen; participants were to type the letters that followed based on alphabetical order, skipping one letter between each entry; for example, if 'A' was presented, participants had to type in the letters 'C', 'E', 'G', and so forth. The delay was manipulated within-subjects, and the alternation between delay and immediate trials was randomised for each participant. After the distractor task, participants provided their rating of the list and then proceeded to the recall task.

Stimuli included 22 six-word lists; six were P1 lists, four were P3/4 (two P3 and two P4), six were P6, and finally six were Control lists. Rotation across participants allowed each list to be presented approximately the same number of times in each of the conditions defined by position and delay condition.

Results and Discussion

An additional exclusion criterion was set for this study: Six participants were excluded because they neglected the distractor task, as their performance ranged from 0% to 4% correct trials (for all other participants' performance was > 85%)

Memory. The recall pattern for the negative items was different for the two delay condition—and in the line with predictions. Both primacy and recency effects were observed in the immediate condition; however, recency effects disappeared after a filled delay (see Figure 2, left panel).

A 3 (position: 1^{st}, $3^{rd}/4^{th}$, and 6^{th}) × 2 (delay: immediate vs. delay) repeated measure ANOVA confirmed the above. The significant main effects of position and delay were qualified by a significant interaction term: There were no recall difference between immediate and delay conditions for negative items presented in the P1 and P3/4 lists (both *p*s > .79); however, for P6 lists, there was a recall advantage for the immediate condition (d = 0.51).

RE. The judgment pattern mirrored the memory results. P1 and P3/4 lists were rated as equally unpleasant in the two delay conditions; on the other hand, P6 lists were rated as more unpleasant in the immediate condition compared to the delay condition (see Figure 2, right panel)

A 3 × 2 repeated measures ANOVA with list type (P1, P3/4 and P6) and delay (immediate vs. delay) confirmed the above: The significant main effects of position and delay were qualified by an interaction. There was no reliable difference in RE between immediate and delay conditions for P1 and P3/4 lists (both *p*s > .19); on the other hand, P6 lists were rated as significantly more pleasant in the delay condition (d = 0.49).

Memory-RE Associations. When the negative item was recalled, RE was lower (*M* = -11.7, *SD* = 9.6) than when it was not recalled (*M* = -5.9, *SD* = 7.0), this pattern being true for both conditions (see Table 4).

A 2 (memory: negative item recalled vs. not recalled) × 2 (delay: immediate vs. delay) within-subjects ANOVA confirmed these observations, as the only significant effect was the main effect of memory.

RE was then analysed depending on the negative item recall output position (see Table 3; note that in this experiment the analyses could not be broken down by delay condition because of the large number of missing values.) A significant main effect of recall position on pleasantness ratings was noted: RE was significantly lower when the negative item was the first or second response than when it was recalled among the last four responses—or not recalled at all (both *d*s > 0.45).

As expected, inserting a filled delay reduced recency effects for recall; recency effects in the delay condition were also significantly reduced for RE. If participants had solely relied on evaluations formed while attending to the stimulus, there would be no reason to expect the observed RE pattern.

Experiment 4

The last experiment aimed to extend the predictions of a memory-based approach to more complex stimuli, namely short stories told through slides with spoken descriptions. We manipulated the position in which a negative slide (a moment within the story) was presented. It was predicted that analyses of memory and summary assessments would again reveal associations between recall and RE.

Method

Participants. A total of 74 participants (36 males) took part in an internet-based experiment. Participants' age ranged from 23 to 68 years (*M* = 44.9, *SD* = 10.7) and they were granted Maximiles® points.

Design, Materials and Procedure. A total of 16 six-slide stories were created, four each for P1, P3/4 (which had a negative slide either in third or fourth position), P6 and Control stories. As negative slides (*M* = 4.2, *SD* = 0.8) were lower in familiarity ratings than neutral slides (*M* = 5.8, *SD* = 1.1), we included a neutral picture which was closely matched for familiarity with the negative slide. The presentation positions of the matched neutral slides mirrored those of the negative pictures. As the position of negative and matched neutral slides was rotated across participants, any potential bias arising from serial position effects should be minimised. Also, the memory analyses will compare the negative slides to the matched neutral pictures only. In this way it was possible to limit the potential confounding effects of familiarity at the (later) retrieval stage.

Negative item rotation meant that each slide (and accompanying narrative) appeared in each possible position (1^{st}, 3^{rd}/4^{th}, and 6^{th}) an equal number of times across participants. Each story was presented in three different versions, without having the overall *gist* changed. Three different versions were created for each of the four Control stories too; this allowed us to check whether re-ordering the slides within a story affected RE. Importantly, the wording which accompanied each slide did not change between the three different versions, so that no difference in either memory and/or RE outputs could be attributed to the wording of narratives.

After attending to a short slideshow and rating its overall pleasantness, participants were asked to complete a memory task. First, participants typed in a brief description of the two moments of the story that came to mind most easily; participants were told that a simple description would suffice. Then, a new screen presented thumbnails of the six slides from the just-presented story, in random order. The participants selected the two pictures that corresponded to the two descriptions they had just typed in; their selections were only used to resolve any potential ambiguity in the interpretation of the descriptions entered in the first phase of the memory task. Two practice trials were provided and included stories which were the same for all the participants and did not include any particularly negative event and/or picture.

Results and Discussion

Memory. Figure 3 (left panel) represents the mean recall proportion as a function of event position and valence. Negative events were better recalled than the neutral ones. Negative events presented in last position attracted the highest recall rate.

A 2 (valence: negative vs. neutral) × 3 (position: P1, P3/4, and P6) repeated measures ANOVA showed that all three effects were significant. The significant interaction showed that the recency effect was stronger for neutral events than for negative events. Most importantly, negative P6 events were better remembered than P3/4 and P1 negative event (both *d*s > 0.28).

Contrary to word lists (Experiment 1), only recency effects were observed for short stories (for similar results with visual stimuli, see Dolenc, Bon, & Repovš, 2013; Hay, Smyth, Hitch, & Horton, 2007). This result is reconcilable with recent findings that suggest that primacy effects in memory are mostly due to rehearsal (e.g., Tan & Ward, 2000). It is reasonable to argue that when the material involved unrelated word lists, rehearsal was more likely than when short stories were involved as the latter promote more relational information processing (e.g., gist extraction).

RE. Preliminary analyses confirmed that changing the order of the slides did not affect RE for Control stories, $F(2, 66) = 1.4$, $p = .23$.

The Control lists were rated as the most pleasant story type. More importantly, recency effects were observed for the pleasantness ratings (see Figure 3, right panel). A repeated measures one-way ANOVA revealed that the main effect of story type was significant and that P6 stories were rated as less pleasant than P1 and P3/4 stories (both *d*s > 0.25).

To confirm that the RE pattern was not influenced by having to perform a memory task, as for Experiment 1 we ran a subsidiary study where only RE ratings were elicited. A total of 113 participants (66 males) took part in an internet-based experiment. Participants' age ranged from 20 to 63 years (*M* = 41.6, *SD* = 12.7) and they were granted Maximiles® points. Recency effects were again observed as P6 stories attracted lower RE than P1 and P3/4 stories (both *d*s > 0.18).

Memory-RE Associations. Table 5 shows that when the negative event was recalled in the memory task, RE for the story was lower (*M* = -23.5, *SD* = 17.2) than when it was not recalled (*M* = -14.7, *SD* = 14.6)—this pattern being constant across story types.

A 2 × 3 within-subjects ANOVA was run with memory (negative event recalled vs. not recalled) and story type (P1, P3/4, and P6) as the factors and revealed a significant main effect of memory that was not qualified by the interaction term, indicating that the association between memory and RE measures was observed regardless of presentation position.

RE was then analysed as a function of narrative accessibility (see Table 3). A significant main effect of negative event output position showed that negative events recalled as first responses were associated with more unpleasant ratings than those recalled as second response (*d* = 0.21).[3]

General Discussion

The above findings provide evidence supporting a memory-based approach to RE. In the experiments involving word-list pleasantness, we manipulated the memorability of a negative item in a two ways, either during encoding (serial position of target item) or after encoding (delay between presentation and recall). In both cases, we correctly predicted changes in RE on the basis of the specific changes in the memorability of the target item. Moreover, list pleasantness ratings were *lower* for the lists where the negative word was recalled—compared to the ratings for the lists where the negative item was not recalled—and *lowest* when it was most easily accessible in memory; the latter result suggests that easily accessible information weighs more heavily in RE and hence contributes to biasing judgments. The basic pattern of results that was obtained for word lists was then reproduced with more complex and easy to integrate materials, namely event-stories. Finally, memory-judgment relationships were observed even when participants did not anticipate a memory task; this result helped to rule out the possibility that participants were artificially induced to rely on the memory representations when providing a RE.

Taken together, the results suggest that participants consulted the episodic record they retained from each sequence in order to assess it in hindsight—and adjusted their judgment based on the information that was most readily available. In effect, two categories of accounts are compatible with our findings. The first suggests that RE is based on memory alone – that is that any assessments that participants make as the experimental events unfold do not contribute to their construction of an overall RE. The second view suggests that RE is the result of on-line impression formation that is adjusted when the summary evaluation is provided; this adjustment would be based on the information that is readily recalled at the time. In other words, this latter view suggests that RE is a form of anchor and adjustment process—a first impression is formed as the information is encountered and this serves as a form of anchor. The latter is then adjusted depending on the information that is most easily available and accessed. The findings we report do not allow us to clearly favour either account –nor were the experiments designed to. However, our results are not compatible with views where memory does not play a role in RE; such views could not predict or explain the findings reported here without putting forward much less parsimonious accounts.

One might be tempted to suggest that the findings are not surprising and that most researchers would expect memory to play a role. However, other judgment literatures strongly argue for the functional independence of memory and judgment; this is the case in a number of instances within work in social cognition (e.g., Betsch, Plessner, Schwieren, & Gütig, 2001). These views hold that stimulus evaluation and encoding into memory are two independent processes, which are called upon depending on the nature of the task at hand. In their influential paper, Hastie and Park (1986) argue that memory-judgment correlations should only arise when participants are not aware of the subsequent judgment task—a condition that does not apply to the present paradigm, as participants were told about the rating task beforehand.

Given the above evidence in favour of the role of memory in biasing RE, one might wonder why correlations between memory and judgment have not always been observed across tasks and domains (for a review, see Hastie & Park, 1986). Although we did not address this question experimentally, we can offer two observations that might shed some light on this issue. One straightforward suggestion is that person-centred judgements (e.g., Asch, 1946) are different to other event sequences. After all, we all have significant experience in judging other people; our judgments will be guided by our theories about other people's motives and our knowledge about how to infer internal states and traits from behaviour (Srull & Wyer, 1989). This being said, assuming there is sufficient commonality between the RE of event sequences and RE of other people, how might we explain that the relationship between memory and judgment is not always reported? First, the present findings suggest that not all the information available in memory has the same influence on retrospective judgment. However, in the social cognition studies where the memory-judgement relation was assessed, all the recalled information is treated as equivalent. For example, together with other studies which investigated memory-judgment correlations (e.g., Lichtenstein & Srull, 1987) Hastie and Park (1986) used ratios as a memory measure. In their Experiment 1, participants heard a recorded 5-min conversation between two people, after which they had to assess the job suitability of one of the two characters and recall the elements they could remember from the conversation. The authors computed a ratio for each participant by dividing the positive arguments recalled (i.e. those supporting candidate suitability) by the total number of arguments remembered: the higher the ratio, the more favourable the memory for the specific candidate. This memory measure was then correlated with the overall job suitability rating the participant provided for the hypothetical character. An implicit assumption underlying the use of this type of ratio is that each argument recalled (and each item on the participants' mind at the time of judgment) has the same weight in the overall evaluation. Such an assumption is disputable in light of the present findings and other theoretical propositions (e.g., Schwarz & Vaughn, 2002; Tversky & Kahneman, 1973) which hold that the ease of recall of a given piece of information influences its effect on evaluations. This would in turn entail that a memory measure (like a ratio) which assigns the same weight to each recalled item may not adequately represent the memory content accessed in order to produce/adjust overall assessments. Relative to previous research that focused on the relationships between memory and judgment (e.g., Anderson & Hubert, 1963; Dreben, Fiske, & Hastie, 1979), the present study utilised more comprehensive memory measures.

A second point relates to the choice of the memory task. Here, participants had to recall the two items that most readily came to mind, and subsequently, any other item they remembered. The major difference with a standard free recall task—which was used in most of the other studies which assessed memory-judgment correlations—is that in the latter participants are asked to recall as many items as they can remember, and this increases the retrieval-related demands of the task. Previous research has highlighted how people seem reluctant to consult all the information they retain about a stimulus in order to judge it; in fact, it seems that they stop their search in memory rather early, as soon as some degree of subjective certainty is achieved (Kitayama & Burnstein, 1989; Schwarz & Vaughn, 2002; see also Higgins, 1996). As evaluations appear based on partial information—on the elements that are most easily retrieved—it is perhaps not so surprising that research calling upon effortful memory tasks, that involve retrieving maximum information, struggled to find correlations between memory and judgment measures (e.g., Anderson & Hubert, 1963; Hastie & Park, 1986; Shedler & Manis, 1986).

Interestingly, different RE patterns were observed for different stimulus types; more specifically, recency effects only were observed for short stories (Experiment 4), whilst Experiment 1 on word-lists displayed both primacy and recency effects. One specific argument we put forward to explain these results relate to retention strategies adopted by the participants: Semantically unrelated word lists may have induced the use of rehearsal strategies more often than short stories, with the latter promoting more relational information processing. In fact, different RE patterns have been observed for different stimulus types and paradigms (e.g., Hogarth & Einhorn, 1992). For instance, Asch (1946) found evidence only for primacy effects in person-centred judgments, whereby early adjectives guided the development of the personality impression; however, in relation to the present memory-based analysis, it is worth noting that in his seminal study Asch (1946) did not measure memory performance for the personality adjectives, and thus it is difficult to draw conclusions about the influence of memory performance on personality impression formation. We acknowledge that different explanations exist in the literature as to when either primacy and/or recency effects are observed in RE (e.g., the nature of the evaluation task; Zauberman, Diehl, & Ariely, 2006). However, for the purpose of the present memory-based analysis, it is worth noting that the pattern of results observed here cannot be accounted for by ‘pure’ on-line judgment and decision-making models. For instance, the Belief-and-Adjustment model (Hogarth & Einhorn, 1992) states that either one of two processing strategies will be adopted when evaluating sequential information. In Step-by-Step (SbS) processing, after each piece of information is encountered the overall assessment of the sequence is reviewed—this leading to a recency effect for judgment, as the last element will have approximately the same weight in the overall evaluation as all the aggregated information that preceded it. Conversely, in End-of-Sequence (EoS) processing, participants perform a single adjustment, as they adjust an initial impression based on early information “by the aggregate impact of the succeeding set of evidence” (Hogarth & Einhorn, 1992; p. 12). EoS processing will therefore lead to primacy effects, as early information has roughly the same impact on summary assessments as all the information following it. The Belief-Adjustment model suggests the choice of strategy is determined by task-related variables (e.g., familiarity of the task, complexity and length of the to-be-evaluated stimuli). Complex and long information series (or evaluations tasks with which participants are rather unfamiliar) would be more likely to lead to SbS processing—and recency effects. On the other hand, simple and short information (or familiar tasks) would be associated with EoS processing—and consequently with primacy biases (Hogarth & Einhorn, 1992). In relation to the current experiments, the important point is that the Belief-and-Adjustment model can explain the recency effects observed in Experiment 4; however, the model would not be able to explain the findings in Experiment 1 as it predicts either primacy or recency effects, whilst both were observed.

The present findings are also in line with other, more recent models that relate memory and judgment processes in order to explain decision-making phenomena. For instance, the Decision-by-Sampling model (DbS; Stewart, Chater, & Brown, 2006) proposes that, when facing a choice, people retrieve from memory (or sample from the decision-making context) instances similar to those at hand (the decision sample). To make a decision, people are thought to rank the target option within the decision sample; the outcome of this comparison will determine the subjective value of the evaluated stimulus. Indeed, previous research has provided support for the notion that people base their judgment on the subset of information that comes to mind when facing a task (e.g., Higgins, 1996). However, the results from the present experiments suggest that, even within this subset, not all information is treated equally. Rather, the influence that specific information has on judgment is proportional to its *relative* accessibility; here, the same negative information had different impact on RE depending on its recall output position: The later the item was recalled, the smaller its weight in RE. In line with this interpretation, and contrary to the DbS assumption that all retrieved stimuli are weighed equally in judgment, empirical evidence suggests that how a specific item in the decision sample differs from the target item may determine its influence on judgment; for example, retrieved values that are less distant from the target item are weighted more heavily in price judgments (e.g., Qian & Brown, 2005).

Recent findings converge on the idea that a memory-based approach can contribute to explaining several judgment and decision-making phenomena. For instance, recently Heit and Hayes (2011) have shown how manipulating the memorability of relevant items successfully predicted participants' performance in property induction tasks. A memory-based approach has been recently shown to explain risky decision-making. Madam, Ludvig, and Spetch (2014) found that participants overweighed the largest gains and losses in a paradigm where they had to make repeated choices between safe and risky options. The authors put forward the 'extreme-outcome' rule whereby the above mentioned biases are explained by the effect that salience exerts on the memorability of rewards. In line with the present series of results, the authors also found that participants tended to recall the extreme outcomes (i.e., the largest gains and losses) early in the recall protocol. Finally, Platzer, Bröder, & Heck, 2014 (see also

Platzer & Bröder, 2012) showed that a memory-based approach can also explain the decision strategies used in multi-attribute decision-making (MADM). In MADM, information elements of different quality need to be integrated in order to reach a decision or make a choice. In their experiments, the authors manipulated the mnemonic accessibility of visual cues by varying their salience. The results showed that when the accessibility of less valid information is increased, participants more often made use of compensatory strategies, whereby all cue information is taken into account in order to take a decision. On the other hand, increasing the accessibility of valid information was more often associated with non-compensatory strategies that are based only on a subset of the available information. Thus, these findings suggest that the interaction between relative accessibility of information and its quality can explain people's decision-making strategies.

In conclusion, the present findings show that a memory-based analysis can predict RE of event sequences even in situations where participants know beforehand that an assessment will be required and in situations where participants do not anticipate a memory test. We hence concur with Montgomery and Unnava (2009) that adopting a memory-based perspective can contribute to a better understanding of the determinants of the RE of temporal sequences. We have argued that while accessibility is most likely not the only contributor to RE, adopting a memory-based analysis can contribute to a more comprehensive explanatory framework. We think such an analysis may be particularly important in applied settings. If it can be established that interfering with the retrieval of negative elements and enhancing memory for more positive events can induce a more positive RE, then the frequency of health-related behaviours—to name one example—can perhaps be increased.

References

Aldrovandi, S., Poirier, M., Kusev, P., Heussen, D., & Ayton, P. (2011). Now I like it, now I don't: Delay effects and retrospective judgment. In Carlson, L., Hoelscher, C., & Shipley, T. F. (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 2866–2871). Austin, TX: Cognitive Science Society.

Anderson, N. H. (1989). Functional memory and on-line attribution. In J. N. Bassili (Ed.), *On-line cognition in person perception* (pp. 175–220). Hillsdale, NJ: Erlbaum.

Anderson, N. H., & Hubert, S. (1963). Effects of concomitant verbal recall on order effects in personality impression formation. *Journal of Verbal Learning and Verbal Behavior, 2,* 379–391.

Ariely, D., & Carmon, Z. (2000) Gestalt characteristics of experiences: the defining features of summarized events. *Journal of Behavioral Decision Making, 13,* 191–201.

Asch, S. E. (1946). Forming impressions of personality. *The Journal of Abnormal and Social Psychology, 41,* 258–290.

Baumgartner, H., Sujan, R., & Padgett, D. (1997). Patterns of affective reactions to advertisements: The integration of moment-to-moment responses into overall judgments. *Journal of Marketing Research, 34,* 219–232.

Betsch, T., Plessner, H., Schwieren, C. & Gütig, R. (2001). I like it but I don't know why: A value-account approach to implicit attitude formation. *Personality and Social Psychology Bulletin, 27,* 242–253.

Bjork, R. A., & Whitten, W. B. (1974). Recency-sensitive retrieval processes in long-term free recall. *Cognitive Psychology, 6,* 173–189.

Bradley, M. M., & Lang, P. J. (1999). Affective norms for English words (ANEW): Instruction manual and affective ratings. *Technical report C-1,* Gainesville, FL. The Center for Research in Psychophysiology, University of Florida.

Chapman, G. B., & Johnson, E. J. (2002). Incorporating the irrelevant: Anchors in judgments of belief and value. In T. Gilovich, D. Griffin, & D. Kahneman, (Eds.), *Heuristic and Biases* (pp.120–138). New York: Cambridge University Press.

Coltheart, M. (1981). The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology, 33A*, 497–505.

Dolenc, B., Bon, J., & Repovš, G. (2013). Serial position and distance effects in visual working memory. *Studia Psychologica, 55,* 67–82.

Dreben, E. K., Fiske, S. T., & Hastie, R. (1979). The independence of evaluative and item information: Impression and recall order effects in behavior-based impression formation. *Journal of Personality and Social Psychology, 37,* 1758–1768.

Fredrickson, B. L. (2000). Extracting meaning from past affective experiences: The importance of peaks, ends, and specific emotions. *Cognition and Emotion, 14,* 577–606.

Fredrickson, B. L., & Kahneman, D. (1993). Duration neglect and in retrospective evaluations of affective episodes. *Journal of Personality and Social Psychology, 65,* 45–55.

Hastie, R., & Park, B. (1986). The relationship between memory and judgment depends on whether the judgment task is memory-based or on-line. *Psychological Review, 93*, 258–268.

Hay, D. C., Smyth, M. M., Hitch, G. J., & Horton, N. J. (2007). Serial position effects in short-term visual memory: A SIMPLE explanation? *Memory & Cognition, 35,* 176–190.

Heit, E., & Hayes, B. K. (2011). Predicting reasoning from memory. *Journal of Experimental Psychology: General, 140*, 76–101.

Higgins, E. T. (1996). Knowledge: Accessibility, applicability, and salience. In E. T. Higgins, & A. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 133–168). New York: Guildford.

Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology, 24*, 1–55.

Kitayama, S., & Burnstein, E. (1989). The relation between opinion and memory: Distinguishing between associative density and structural equality. In J. N. Bassili (Ed.), *On-line cognition in person perception* (pp. 91–122). Hillsdale, NJ: Erlbaum.

Kucera, H., & Francis, W. N. (1967). Computational Analysis of Present-Day American English. Brown University Press, Providence R.I.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1999). *The international affective pictures system (IAPS). Technical manual and affective ratings*. Gainsville, FL: University of Florida.

Langer, T., Sarin, R. K., & Weber, M. (2005). The retrospective evaluation of payment sequences: Duration neglect and Peak-and-End effects. *Journal of Economic Behavior and Organization, 58*, 157–175.

Lichtenstein, M., & Srull, T. K (1987). Processing objectives as a determinant of the relationship between recall and judgment. *Journal of Experimental Social Psychology, 23*, 93–118.

Loewenstein, G., & Prelec, D. (1993). Preferences over sequences of outcomes. *Psychological Review, 100*, 91–108.

Madam, C. R., Ludvig, E. A., & Spetch, M. L. (2014). Remembering the best and worst of times: Memories for extreme outcomes bias risky decisions. *Psychonomic Bulletin & Review, 21*, 629–636.

Montgomery, N. V., & Unnava, H. R. (2009). Temporal sequence effects: A memory framework. *Journal of Consumer Research, 36,* 83–92.

Moser, D. V. (1992). Does memory affect judgment? Self-generated versus recall memory measures. *Journal of Personality and Social Psychology, 62,* 555–563.

Platzer, C., & Bröder, A. (2012). Most people do not ignore salient invalid cues in memory-based decisions. *Psychonomic Bulletin & Review, 19,* 654-661.

Platzer, C., Bröder, A., & Heck, D. W. (2014). Deciding with the eye: how the visually manipulated accessibility of information in memory influences decision behavior. *Memory & Cognition, 42,* 595–608.

Qian, J., & Brown, G. D. A. (2005). Similarity-based sampling: Testing a model of price psychophysics. In Bara, B. G., Barsalou, L., & Bucciarelli, M. (Eds.), *Proceedings of the 27th Annual Conference of the Cognitive Science Society* (pp. 1785–1790). Mahwah, NJ: Erlbaum.

Redelmeier, D. A., & Kahneman, D. (1996). Patients' memories of painful medical procedures: Real-time and retrospective evaluations of two minimally invasive procedures. *Pain, 66,* 3–8.

Schafer, J. L., & Olsen, M. K. (1998). Multiple imputation for multivariate missing-data problems: A data analyst's perspective. *Multivariate Behavioral Research, 33,* 545–571.

Shedler, J., & Manis, M. (1986). Can the availability heuristic explain vividness effects? *Journal of Personality and Social Psychology, 51,* 26–36.

Schwarz, N., & Vaughn, L. A. (2002). The availability heuristic revisited: Ease of recall and content of recall as distinct sources of information. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristic and Biases* (pp.103–119). New York: Cambridge University Press.

Srull, T. K., & Wyer, R. S. (1989). Person memory and judgment. *Psychological Review, 96*, 58–83.

Stewart, N., Chater, N., & Brown, G. D. A. (2006). Decision by Sampling. *Cognitive Psychology, 53*, 1-26.

Tan, L., & Ward, G. (2000). A recency-based account of the primacy effect in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1589–1625.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 5*, 207–232.

Weber, E. U., & Johnson, E. J. (2009). Mindful judgment and decision making. *Annual Review of Psychology, 60*, 53-85.

Zauberman, G., Diehl, K., & Ariely, D. (2006). Hedonic versus informational evaluations: Task dependent preferences for sequences of outcomes. *Journal of Behavioral Decision Making, 19*, 191–211.

Footnotes

[1] This analysis yielded missing values (ranging from 2.8% to 10.4% across experiments). Missing values were replaced using different methods, including mean substitution by subject, grand mean, and Expectation-Maximization algorithm (Schafer & Olsen, 1998). As all the analyses returned the same results in all cases, we will be reporting the data obtained via mean by subject substitution. Also, here and throughout the paper all the missing values were missing completely at random as the MCR Little's test was never significant (all ps > .15).

[2] Interestingly, the pleasantness ratings for those lists where the negative item was presented but not recalled were very close to zero. This finding would suggest that, as memory task demands are reduced to zero, a pure memory-driven approach can best explain retrospective judgment. However, the one-trial nature of this experiment would suggest caution in interpreting the results—further studies need to be conducted to assess the reliability of this finding.

[3] In order to rule out the possibility that demand characteristics artificially inflated the findings about the relationship between memory and judgment, additional considerations and analyses are offered here. First, it is important to notice that in Experiment 2, where participants performed a surprise memory task and thus could not strategically adjust their evaluation strategies, the observed relationship between memory and judgment is similar to that observed in the other experiments—in fact, it is even stronger. Second, we would argue that, due to the repeated trials, participants expected the summary evaluations to be prompted after the presentation of the stimuli. As such, according to the analysis of Hastie and Park (1986), this paradigm should have promoted online judgment (as opposed to memory-based) strategies, which, in turn, should have actually hindered any relationship between memory and judgment. Third, we have run additional analyses that explored the relationship between

memory and judgment when only the first few trials are considered. In Experiment 1, when only the first trial is considered, the average corrected pleasantness rating when the negative item was not recalled was -13.0 (*SD* = 11.7); the corrected judgment for those trials where the negative item was recalled was -16.3 (*SD* = 17.4; $t(34) = 0.5$, $p = .613$). We acknowledge that the difference was not significant, although this result was probably due to the variability in the judgment scores and the insufficient power to detect the underlying effect. Indeed, when the first two trials were included in the analysis, the difference approached significance (p = .132), which was reached when the first three trials were analysed ($p = .020$). The pattern of results was similar for Experiment 3: When only the first trial was analysed, the average corrected pleasantness rating when the negative item was not recalled as -6.9 (*SD* = 12.5); the corrected judgment for those trials when the negative item was recalled was -9.6 (*SD* = 15.5; $t(70) = 0.7$, $p = .494$). When the second trial was included in the analysis, the above difference became significant ($p = .035$). A similar pattern is observed when the analysis was split by delay condition. Finally, in Experiment 4 the results were more clear cut as showed a significant effect when only the first trial was included in the analysis: The average corrected pleasantness rating when the negative item was not recalled was -6.7 (*SD* = 13.4); the corrected judgment for those trials when the negative item was recalled was -22.9 (*SD* = 18.4; $t(54) = 3.7$, $p < .01$).

Table 1 *ANOVA statistics for all analyses*

Analysis	Experiment	Effect	df	*F*	*p*	η_p^2
Memory	Experiment 1	Position	1.7, 59.0	18.7	<.001	.35
		Valence	1, 35	12.6	<.01	.27
		Position*Valence	2, 70	2.3	.11	
	Experiment 3	Position	1.8, 127.6	28.8	<.001	.29
		Delay	1, 71	10.1	<.01	.12
		Position*Delay	2, 142	3.5	<.05	.05
	Experiment 4	Position	1.8, 123.3	32.7	<.001	.33
		Valence	1, 68	94.6	<.001	.58
		Position*Valence	2, 136	9.7	<.001	.13
RE	Experiment 1	List type	3, 105	37.1	<.001	.51
	Experiment 3	List type	1.8, 127.1	3.3	<.05	.04
		Delay	1, 71	7.7	<.01	.10
		List type*Delay	2, 142	7.4	<.001	.09
	Experiment 4	List type	1.8, 120.6	87.6	<.001	.56
Memory-RE	Experiment 1	Memory	1, 35	16.2	<.001	.32
		Memory*List type	1.9, 69.1	1.4	.12	
		Recall position	1.7, 58.3	22.0	<.001	.39
	Experiment 2	Recall position	2, 104	5.0	<.01	.09
	Experiment 3	Memory	1, 68	40.6	<.001	.37
		Delay	1, 68	2.6	.11	
		Memory*Delay	1, 68	0.9	.34	
		Recall position	2, 128	15.8	<.001	.20
	Experiment 4	Memory	1, 68	30.0	<.001	.31
		Memory*List type	1.7, 116.9	2.4	.07	
		Recall position	1.6, 109.9	18.7	<.001	.22

Note. When degrees of freedom are not presented as integers, it is because they are adjusted according to the Greenhouse-Geisser correction due to violations of the sphericity assumption.

Table 2

Mean corrected pleasantness ratings (J') and standard deviations (SD) as a function of list type and negative item being recalled or not (Experiment 1)

		List type		
		P1	P3/4	P6
Was the negative item recalled?				
No	*M*	-7.7	-7.6	-11.1
	SD	(12.1)	(10.2)	(12.5)
Yes	*M*	-18.3	-14.4	-15.6
	SD	(13.0)	(12.1)	(11.4)

Table 3

Mean corrected pleasantness ratings (J') and standard deviations (SD) as a function of negative item negative accessibility, assessed via its recall output position (see text for further details)

		Negative item recall output position		
		Not recalled	Recalled early	Recalled later
Experiment 1	*M*	-7.6	-19.6	-11.3
	SD	(8.5)	(12.5)	(9.0)
Experiment 2	*M*	0.3	-13.0	-3.7
	SD	(20.7)	(21.3)	(14.7)
(between-subjects)	*n*	38	49	20
Experiment 3	*M*	-6.9	-13.7	-8.6
	SD	(7.0)	(10.9)	(11.5)
Experiment 4	*M*	-14.7	-24.6	-20.9
	SD	(14.6)	(17.6)	(17.5)

Table 4

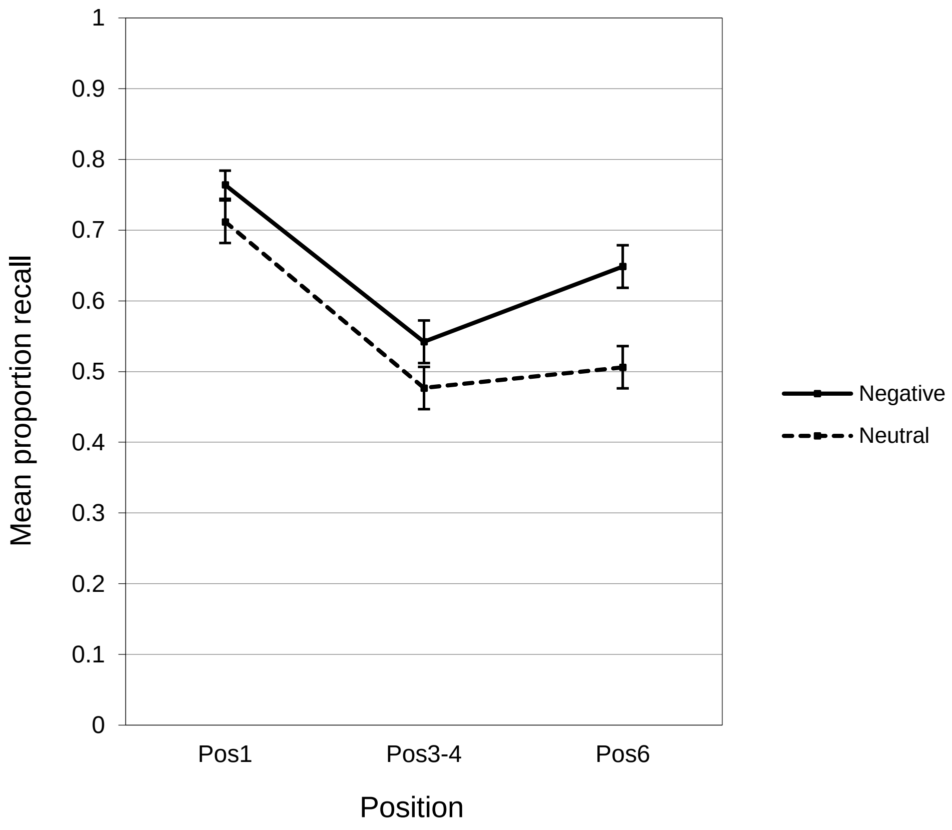
Mean corrected pleasantness ratings (J') and standard deviations (SD) as a function of list type and negative item being recalled or not (Experiment 3)

		Delay condition	
		Immediate	Delay
Was the negative item recalled?			
No	*M*	-7.1	-4.7
	SD	(10.2)	(7.2)
Yes	*M*	-12.2	-11.2
	SD	(10.3)	(10.9)

Table 5

Mean corrected pleasantness ratings (J') and standard deviations (SD) as a function of story type and the negative event being recalled or not (Experiment 4)

		Story type		
		P1	P3/4	P6
Was the negative event recalled?				
No	*M*	-14.6	-19.1	-20.3
	SD	(15.4)	(15.8)	(16.0)
Yes	*M*	-22.8	-22.7	-24.6
	SD	(18.3)	(17.8)	(17.9)



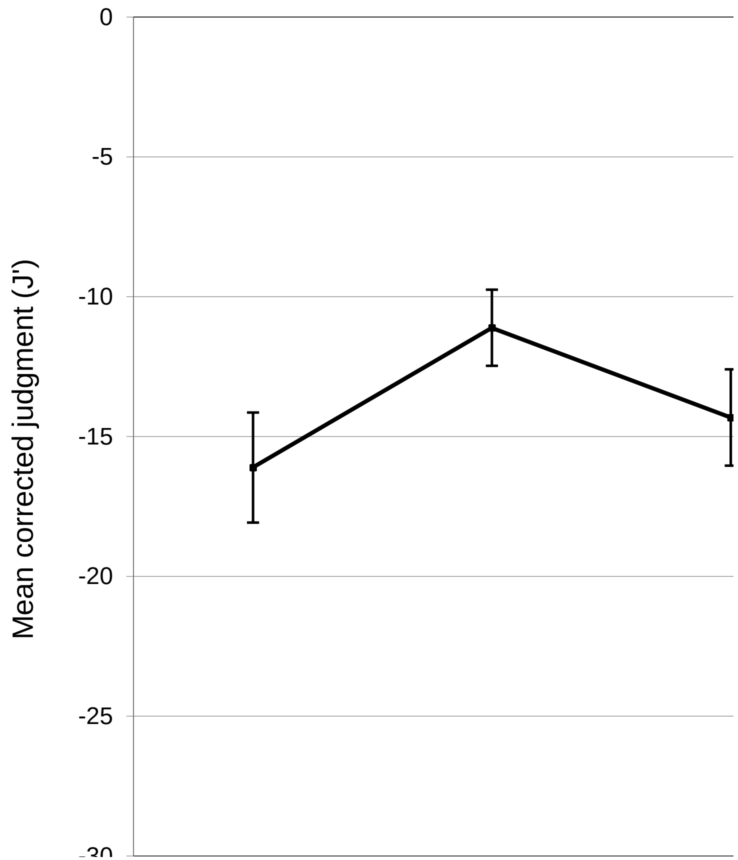


Figure 1. Experiment 1 mean results for (a) Memory (left panel): mean recall as a function of word position and valence and (b) RE (right panel): mean pleasantness ratings as a function of list type. Error bars indicate *SEM*.

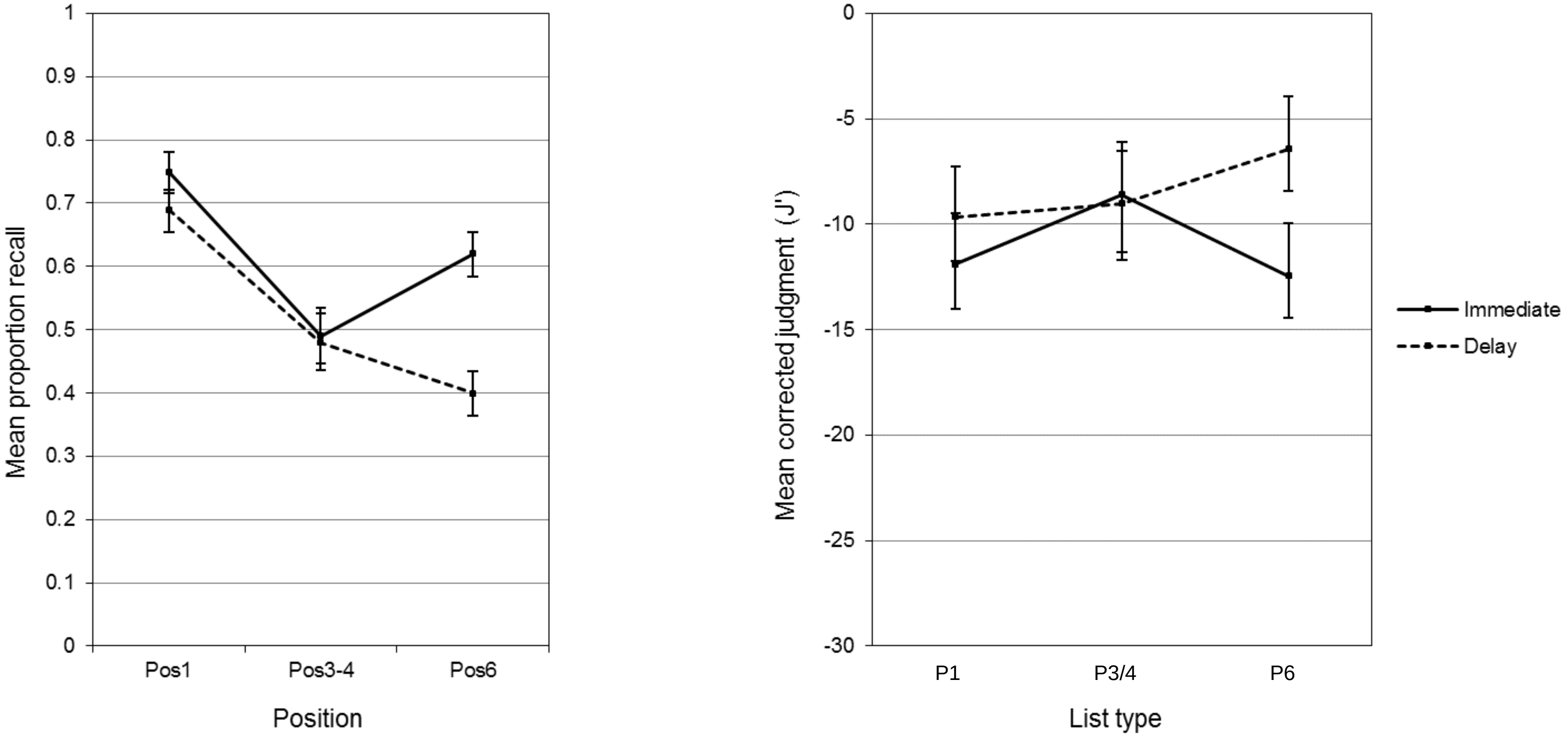


Figure 2. Mean proportion recall for negative items as a function of word position and delay condition (left panel) and mean pleasantness ratings as a function of list type and delay condition (right panel) Error bars represent *SEM* (Experiment 3).

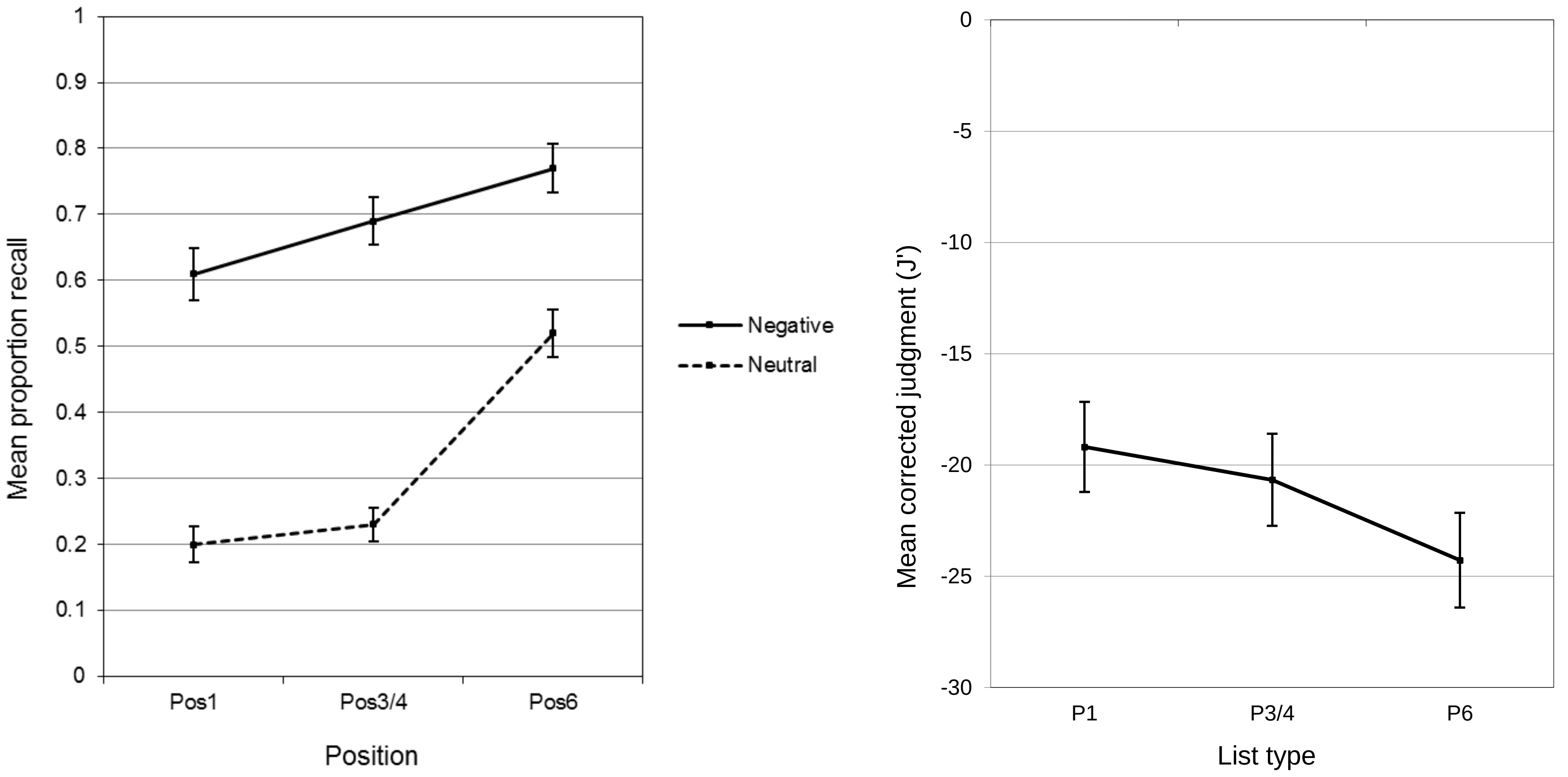


Figure 3. Mean recall proportion as a function of event position and valence (left panel) and mean pleasantness ratings as a function of story type (right panel). Error bars indicate *SEM* (Experiment 4).

Appendix

Word lists used for Experiment 1.

List no.	Items
1	victim, basket, cabinet, stomach, rabbit, avenue
2	rape, lawn, chair, hay, foot, coin
3	devil, bench, tool, nectar, queen, trunk
4	slave, paint, sphere, dawn, hammer, seat
5	riot, lamb, ink, spray, custom, gender
6	pain, cook, farm, dress, moral, hat
7	accident, cow, corridor, machine, thermometer, humane
8	afraid, column, smooth, vision, journal, phase
9	anger, dream, quiet, tune, tower, bottle
10	tragedy, passage, wonder, clothing, opinion, lantern
11	horror, priest, statue, fabric, flower, breeze
12	hate, item, nice, doll, silk, tree
13	terrible, contents, sentiment, adult, industry, stove
14	cancer, pencil, locker, tender, ankle, circle
15	agony, obey, hawk, quart, saint, jelly
16	danger, wine, daylight, poetry, paper, humble
17	punishment, bandage, intellect, village, indifferent, umbrella
18	destroy, serious, prairie, garment, nonsense, teacher
19	burn, cord, iron, tank, tidy, fork
20	disaster, coast, orchestra, appliance, modest, kerchief
21	thief, bowl, vest, swamp, runner, nun
22	hurt, coarse, unit, clock, save, pet
23	trouble, natural, building, manner, quality, window
24	slap, kettle, poster, tennis, trumpet, scissors
25	candy, cannon, banner, lake, yellow, stool
26	bereavement, barrel, elbow, milk, nonchalant, grass
27	cat, autumn, cottage, frog, glacier, salad
28	owl, fur, green, horse, key, letter
29	engine, kerosene, dove, honey, egg, watch
30	chin, crown, metal, plant, radio, rock
31	fish, bar, movie, mystic, truck, butter
32	flag, person, rattle, river, theory, violin

Notes. Normative values about word valence, arousal, familiarity and frequency can be obtained from the first author. The negative item is always listed as the first item. Lists 25-32 are Control lists.