Memory and Judgment Bias in Retrospective Evaluations

Silvio Aldrovandi

Thesis submitted in fulfilment

of the requirements for the degree of

Doctor of Philosophy

Department of Psychology
City University, London

May 2009

Table of contents

List of tables

List of figures

Acknowledgments

I would like to express my deepest gratitude to my supervisor, Marie Poirier, for her guidance throughout the PhD. Marie has led my curiosity into practicable research paths – and helped me not to get lost. She has also (patiently!) reviewed my manuscripts and provided me with invaluable feedback. Marie has been a great mentor, both at academic and personal levels.

This thesis would have not been possible without the everyday encouragement, support, and love of my wife, Sanjna. We surpass together the difficulties, and together we celebrate the achievements.

I want to say a big thank you to my parents, because they unconditionally support all my choices and they always believe in me.

Thank you to all the friends and colleagues on the 4th and 5th floor of the Social Science building. Especially to Daniel Heussen, for the fruitful discussions about my research, for "dragging me" into teaching and for the feedback on my write-up. To Lina Daouk, whose social initiatives brought together a wonderful group, which made the PhD fun, too. To Sebastian Gaigg who provided me with suggestions about my research. And to Emma Walsh, with whom we discuss our progress and push each other into respecting deadlines.

For the productive discussions on the work reported in the present thesis I would like to thank the new – and "old" – members of the Memory Research Unit: Kiki, James, Camilla, and Jonathan. For the same reason, a big thank you goes also to Petko Kusev, Peter Ayton, the Memory & Cognition Research Group at City University, and the London Judgment and Decision Making group at UCL.

I am very grateful to Peter Barr for his help with the programming and Web-publishing of Experiments 7–10, and to Neha Malhotra, who helped me out in collecting the data for Experiment 3.

During the PhD I very much enjoyed the teaching experience I gathered – for this reason I want to thank Daniel Heussen, James Hampton, Peter Martin, Monica Magadi, and Paul Flaxman.

I want also to thank City University, who welcomed me and provided me with the financial support and a thriving academic environment – and all the participants who took part in my studies and made this research possible.

Thank you to Angelo, Marcella, the ballotta in Bologna, Maya and Mohun, Sam, Christoph and Mirjam, for being close to me and supporting me despite the geographical distance. Finally, I am grateful for the teachings of Giancarlo de Petrillo and Duane Chapman.

Declaration

I grant powers of discretion to the University Librarian to allow this thesis to be copied in whole or in part without further reference to me. This permission covers only single copies made for study purposes, subject to normal conditions of acknowledgement.

Abstract

Previous research has shown that people are biased when providing summary assessments about past experiences. Retrospective evaluations seem based on specific moments within the to-be-assessed episode (e.g., Fredrickson & Kahneman, 1993; Hogarth & Einhorn, 1992). Here, drawing on some of the parallelisms between memory phenomena and judgment and decision making biases (e.g., distinctiveness; Hunt, 2006), a memory-based approach is outlined and explored. By doing so, the research also addressed the broader debate surrounding the relationships between memory and judgment, fuelled by conflicting results about memory–judgment correlations (e.g., Hastie & Park, 1986). In the first set of studies, participants recalled lists of words, after having assessed each list for pleasantness. The results showed clear associations between memory and judgment, which appeared moderated by the cognitive demands associated with the memory task. Retrospective evaluations were predicted by both the content of recall and the ease with which distinctive information was brought to mind (Schwarz, 1998). The nature of these associations was further investigated in the second set of studies. By hindering the memorability of negative information, it was possible to reduce its impact on retrospective assessments; this finding can easily be interpreted if one assumes a causal relationship between memory and judgment. The last study corroborated the memory-based approach with more cohesive stimuli – namely short stories told through slideshows. These results also suggested that memory-judgment relationships are vulnerable to experiment-related factors; they are easily hindered if the way memory is probed is not aligned with the information that underlies the judgment task. Overall, the findings suggest that, in order to assess an event in retrospect, people rely on their memory but in a way that is biased by the relative availability of certain features of the event. These results are at odds with the predictions of on-line judgment theories (e.g., Anderson, 1989), which postulate functional independence between memory and judgment. The theoretical and practical implications of this work are discussed and future directions for research on the role of memory in retrospective evaluations are suggested.

Chapter 1:

General Introduction

1.1. Overview of the General Introduction

The General Introduction is divided into main sections. First, retrospective evaluations are defined – and their importance in everyday cognition highlighted. Then, the social cognition and decision-making literatures which investigated summary assessments are reviewed, with a focus on the most frequently observed biases.

By outlining some of the parallels between memory and judgment phenomena, it is argued that memory functioning can account for the array of judgment biases in retrospective evaluations. I review the literature on the relationships between memory and judgment and justify why it is important to gather further – and more direct – evidence on the role of memory in retrospective judgments.

An overview of the present thesis completes the present chapter. Finally, I briefly introduce each of the following four chapters, outlining the rationale and objectives of each.

1.2. Biases in retrospective evaluations

1.2.1. Why is it important to study RE?

The complexity of evaluating the past

Think back to the last film you watched at the cinema. How was it? Did you like it? How *much* did you like it? The responses people provide to these sorts of questions, despite often being provided effortlessly, conceal an important and complex aspect of everyday cognition. Films – as most of the episodes in our

lives – are experiences that unfold over time. Undoubtedly, the perceived quality of the film was not constant from its beginning to its end: Some scenes were more effective than others, some parts more interesting. How do people provide *retrospective evaluations*? That is, how do they answer questions about the *overall* quality of extended episodes?

When people assess past events, they tend to provide *summary assessments* of the episode under consideration – instead of moment-by-moment accounts about how they felt while the event was unfolding (Fredrickson, 2000; Kahneman, 2000b). It follows that people must somehow take into consideration the quality of the different moments which defined the episode in order to evaluate the latter in hindsight. When this is done, a global assessment is produced and it represents a unitary and coherent judgment – about an event that may not have been so.

Ubiquity of retrospective evaluations

Summary assessments are not only about the quality of extended episodes or experiences. Often people evaluate a target stimulus on the basis of the information they acquire about it. Inevitably, the pieces of information we acquire about something are often different in quality and importance; nevertheless, we are able to provide a global judgment – a coherent answer which represents potentially incoherent dimensions of the information. Consider the task participants were faced with in the seminal work of Solomon Asch on personality impression formation (1946; cited in Zauberman, Diehl, & Ariely, 2006). Asch (1946) asked participants to provide a global rating of likeableness of a hypothetical character, which was described through a series of adjectives.

The person was described as *envious*, *stubborn*, *critical*, *impulsive*, *industrious*, and *intelligent*. Clearly, participants had to summarise divergent information (or inputs of different quality) in order to complete the task – similarly to what is done when rating how much a film is liked.

Evaluating information that is presented sequentially or events that unfolded over time is a common cognitive task that people have to perform everyday, and it is not surprising that many studies have investigated *how* people provide summary assessments. Retrospective evaluations have been studied in multiple domains, including: consumer behaviour (Russo, Meloy, & Medvec, 1998), economics (Betsch, Plessner, Schwieren, & Gütig, 2001; Langer, Sarin, & Weber, 2005), medical settings (Redelmeier & Kahneman, 1996; Stone, Schwartz, Broderick, & Shiffman, 2005), jury simulations (Reyes, Thompson, & Bower, 1980), personality attribution (Asch, 1946; Lichtenstein & Srull, 1987), perception (Ariely, 2001), and politics (Lodge, McGraw, & Stroh, 1989).

Importance of retrospective evaluations

Understanding how people evaluate the past has important implications. The work by Kahneman and colleagues suggests that how people summarise experiences in hindsight drives the choices they will make in the future (Kahneman, 2000a; 2000b; Kahneman & Thaler, 2006; Kahneman, Fredrickson, Schreiber, & Redelmeier, 1993; Kahneman, Wakker, & Sarin, 1997; see also Read, 2004). Retrospective evaluations appear to be an important input into decisions to repeat (or not repeat) past experiences. For example, a survey by Baines, To, and Wall (1990), revealed that patients' memories of unpleasant medical procedures influenced their decisions about future treatment choices; for

example, about 20% of women who did not attend a routine mammogram screening mentioned the remembered pain of previous screenings as the reason for their decision (see also Elwood, McNoe, Smith, Bandaranavake, & Doyle, 1998). More strikingly, about 40% of patients who survived a cardiac arrest decided not to undergo future revival procedures – and again the remembered discomfort of the treatment appears to have been a strong determinant of their choice (Bedell, Delbanco, Cook, & Epstein, 1983).

Most importantly, Kahneman and colleagues (see also Oishi & Sullivan, 2006) have argued that the way people summarise the past (the *remembered* utility) can be a better predictor of future choices (*decision* utility) than what they actually experienced (*experienced* utility). Let me illustrate by briefly describing one of their studies (Kahneman et al., 1993). Participants volunteered to immerse both hands (at different times) in cold water – experiences which were uncomfortable in nature. A "short" immersion lasted for 60 seconds and the water was at a temperature of 14°C. The "long" immersion was the same as the short one, with the only difference of an additional 30 seconds at a temperature of 15°C – i.e. warmer, but still at uncomfortable levels. Thus, participants experienced more overall discomfort during the long immersion, as it lasted for a longer time; that is, its *experienced* utility was lower. However, when asked to rate the two immersions on a series of dimensions (e.g., "overall discomfort", p. 402) participants disliked the long immersion the least; that is, they rated its *remembered* utility as higher. This finding already exemplifies how the summarised quality of an episode can differ from its actual quality – point which I will discuss in further detail in the next sections. Ultimately, when participants were asked to choose which of the two immersions they would repeat – that is,

they were asked to provide the *decision* utility – they preferred the long immersion. This choice was therefore predicted by how participants summarised the experience, rather than its actual overall quality. As participants' choice led them to be exposed for a second time to the longer immersion (i.e. characterised by a higher total amount of discomfort) the authors argue that *remembered* utility can push people into counterintuitive and potentially maladaptive choices (Kahneman et al., 1993).

In summary, it would appear that it is not what we experience that drives our choices for the future – it is how we summarise it in hindsight. As we will discuss further in the next paragraph, this is a very important distinction: Very often the way people summarise experiences departs significantly from what they actually experienced.

Are people biased when evaluating the past?

When summarising past episodes (or information presented sequentially), do people equally weigh each moment of the experience (or piece of information within the sequence)? Or, is the perceived quality of an outcome (its utility) equal to the summed quality of its components (e.g., Samuelson, 1937; cited in Stewart, Chater, & Brown, 2006)? The following two sections will cover the research that attempted to answer the above questions. In order to do so, two separate lines of research need to be reviewed; one is from the judgment and decision-making literature while the other is taken from the information integration and social cognition literatures.

1.2.2. Biases in judgment and decision making

The "Peak-End" rule

Recent work by Kahneman and colleagues (e.g., Fredrickson & Kahneman, 1993; Schreiber & Kahneman, 2000; Varey & Kahneman, 1992) has suggested that when people summarise the quality of an event in hindsight, they are largely influenced by the quality of two of its defining moments "in an almost rule-like fashion" (Fredrickson, 2000; p. 579). First, the "Peak": this is the moment during which the *intensity* of the experience is at the highest; at this moment, the quality departure from the average quality of the *whole event* is the maximum. Second, the "End": this is the quality of the final moments of the experience. By averaging the quality (utility) experienced during these two moments, it is possible to successfully predict the summary assessments for the experience *as a whole* (Fredrickson & Kahneman, 1993); how pleasant or unpleasant a specific episode is on average does not seem to matter as much (but see Cojuharenco & Ryvkin, 2008).

In a study ran by Varey and Kahneman (1992), participants were presented with a series of ratings indicating the discomfort that a hypothetical character experienced during activities like standing in an uncomfortable position and being exposed to loud noises. The participants' task was to rate the character's global discomfort. Ninety-four percent of the variance in the global assessments was accounted for by the average of the "Peak" – the most extreme discomfort reported by the character – and the "End", i.e. the final ratings (Fredrickson & Kahneman, 1993; Schreiber & Kahneman, 2000). In their discussion, the authors conclude that "the results showed that extended aversive

episodes are not evaluated by integrating the disutility of successive intervals" (Varey & Kahneman, 1992; p. 181).

Similar conclusions were reached in a study where participants directly experienced the to-be-assessed events – namely computer-generated sounds (Schreiber & Kahneman, 2000). First, the authors asked participants to continuously appraise their annoyance while being presented with unpleasant sounds of varying intensity. As the reported annoyance closely tracked the sounds intensity, the authors could use the latter as a measure of the discomfort experienced by participants at any moment. In the next experiments, the authors manipulated the intensity and duration of sequence of annoying sounds; the participants' task was to rate the overall (un)pleasantness of each sequence of sounds, once it finished. The results supported the authors' claim that "…peak and end are good predictors of remembered utility" (p. 40). Most notably, a simple average of the intensity at the "Peak" moment and at the "End" better predicted participants' summary assessments than the total or the average intensity of the sound sequences.

Peak-End phenomena have been observed in many laboratory-based studies. Some of these studies involved experiences relatively aversive in nature, like discomfort from immerging a hand in cold water (Kahneman et al., 1993), from annoying sounds (Västfjäll, 2004), and pain from heat or mechanical pressure (Ariely, 1998). In other studies, participants had to evaluate relatively more pleasant experiences, such as TV advertisement (Baumgartner, Sujan, & Padgett, 1997), movie clips (Fredrickson & Kahneman, 1993), musical selections (Rozin, Rozin, & Goldberg, 2004), monetary sequences (Langer et al., 2005), and material goods (Do, Rupert, & Wolford, 2008). Moreover, Peak-End effects

have been observed in real-life situations, including medical (Oliver, 2008; Redelmeier & Kahneman, 1996; Stone, Broderick, Kaell, DelesPaul, & Porter, 2000) and job settings (Clark & Georgellis, 2004). The range of experiences where the Peak-End rule seems to apply is rather large; this includes episodes that were either short or prolonged (Clark & Georgellis, 2004; Schreiber & Kahneman, 2000) and finally, episodes that were continuous or formed of a sequence of discrete stimuli (Kahneman et al., 1993; Langer et al., 2005).

Duration neglect and the violation of the temporal monotonicity principle

A direct – and experimentally validated – consequence of "Peak" and "End" being overweighed in summary assessments is that the principle of "temporal monotonicity" (cf. Fredrickson, 2000) is often violated. This principle states that adding a period of negative affect to an ongoing experience should make the overall experience worse; on the other hand, extending the experience with positive affect should instead make the overall experience better. This would follow if retrospective evaluation was based on an averaged impression of an experience; however, if the main determinants of remembered utility are the Peak and End of a sequence, the overall duration of an event should not be an important factor in is appreciation in hindsight. In an oft-cited clinical study, Redelmeier, Katz and Kahneman (2003) showed that adding an extra period of pain can actually improve the evaluation of a painful medical screening procedure. Patients undergoing colonoscopy were randomly assigned to either a control group, who underwent the standard procedure, or an experimental group. In the latter condition, the procedure was extended by leaving the apparatus in place for an average two minutes after the clinic examination was completed.

This added experience was mildly uncomfortable – but less painful than the preceding moments. That is, the period of added pain represented a noticeable improvement. Indeed, as patients in both groups were given the possibility of rating their current pain in a real-time mode, it was possible to determine that the experimental group rated the final moments as less painful. The main results confirmed the authors' predictions. First, the patients in the experimental group rated the colonoscopy *as a whole* as less painful than patients in the control group. The authors argued that this result was due to the "End" being largely overweighed in summary assessments. Indeed, despite the overall amount of experienced pain being greater for the experimental group, the final moments were perceived as less painful. Second, confirming the link between retrospective evaluations and decisions about the future, the patients from the experimental group were slightly—but significantly—more likely to attend future recommended screenings.

As people seem to summarise the quality of past events on the basis of a few of their defining moments ("snapshots"; cf. Fredrickson & Kahneman, 1993), it follows that more *holistic* features are often neglected. For instance, it seems that people tend to disregard the duration of the events when evaluating them in hindsight (e.g., Diener, Wirtz, & Oishi, 2001; Hands & Avons, 2001; Rode, Rozin, & Durlach, 2007; Rozin et al., 2004). Closely related to the violation of the temporal monotonicity principle, duration neglect may lead people to puzzling choices. Going back to a study on pain perception, Kahneman et al. (1993) showed that, despite being able to judge correctly the duration of two pain-inducing experimental trials, participants chose the one which lasted longer (see also Redelmeier & Kahneman, 1996).

Limits of the "Peak-End" rule: When holistic features matter

Although there is considerable empirical evidence that supports the "Peak-End" rule, several authors have questioned its generality. For example, according to Ariely and colleagues, people do not always neglect event duration when assessing experiences in hindsight – especially when it is made an explicit and salient attribute of the episode (see Ariely, 1998; Ariely & Loewenstein, 2000; Ariely, Kahneman, & Loewenstein, 2000; Ariely & Zakay, 2001). In addition, Liersch and McKenzie (2009) have recently suggested that duration neglect may be moderated by situational factors – for instance, the way the information is presented to participants. The authors, similarly to Varey and Kahneman (1992), asked participants to rate the overall discomfort of a hypothetical character who was engaged in uncomfortable activities and had rated the moment-by-moment the discomfort while experiencing them. When the momentary ratings were presented via a sequence of numerical values, participants displayed duration neglect in their overall assessments. However, when the information was presented via graphs, duration did affect participants' retrospective evaluations – as the rated unpleasantness of the experiences increased along with their duration.

Some authors have argued that, beyond the experience duration, other holistic features of an event may be important, too (e.g., Ariely & Carmon, 2000). For instance, some studies highlighted how the episode *trend* – the overall rate and direction of hedonic change – affects summary assessments (Ariely & Zauberman, 2000; 2003; Diehl & Zauberman, 2005; Hsee, Abelson, & Salovey, 1991; Jochen & Cropanzano, 2007; Loewenstein & Prelec, 1993). Simply put, it

seems that experiences that improved over time were preferred to experiences whose quality decreased in the last moments (Loewenstein & Prelec, 1993). As the trend of an experience must involve some processing of *relational* information, it might be an over-generalisation to assume that retrospective evaluations are always based on 'stills' or moments of a remembered experience.

Another line of evidence suggests that specific task conditions may moderate "Peak", "End" – and even "trend" – biases in retrospective evaluations. Ariely (1998) had participants experience short and unpleasant episodes, i.e. exposure to heat or mechanical pressure. In one condition, participants were asked to rate the intensity of the stimuli while experiencing them ("On-line +" condition; p. 31) and to rate their overall discomfort. In the control condition ("On-line –"; p. 31), participants only rated the experience once it ended. The results clearly showed that continuously reporting the intensity of the stimuli reduced the bias in the overall assessments. Indeed, the retrospective evaluations provided by the participants in the On-line + condition were less affected by the "Peak", "End" and "trend" features of the episodes – and reflected more closely the average quality of the experience (see also Ariely & Zauberman, 2000).

Conclusions about hedonic episodes

Despite the above mentioned caveats, a general conclusion can be safely drawn from the literature reviewed so far: When assessing the past, people seem to overweigh specific parts of the experience – may they be the "Peak", the "End" or the more general "trend". Indeed, it has been shown that the parts of the trend of an experience that matter the most are those at the end of the experience (Ariely & Zauberman, 2000; 2003); hence, the influence of the trend on

retrospective evaluations can be reformulated as the influence of the final part of the experience – rather than its final moments as predicted by the "Peak-End" rule.

Thus, the answer from the judgment and decision-making literature to the question previously introduced– whether people are biased when evaluating the past – seems to be a resounding "yes". Specific moments within the to-be-assessed hedonic event are largely overweighed. Situational factors unrelated to the experience may moderate the bias in retrospective evaluations (e.g., personality disposition; Barrett, 1997). Nevertheless, it can be safely put forward that summary assessments are largely influenced by key moments within the to-be-assessed outcome.

Next, we will explore whether the above – general – conclusion still holds for a different branch of research, which investigated information integration in hindsight judgment.

1.2.3. Order effects in information integration

"First takes all" and "save the last dance for me"

Evaluating streams of information that have been provided in a sequential manner is another frequent activity that people are faced with. When talking to a sales assistant about selecting a laptop to purchase, it is likely each laptop will be associated with a list of features, presented in sequence for each potential 'candidate'; the assessment of this sequence of features will guide our evaluation of the laptops, and consequently the choice about which one to buy.

The extensive work by Russo and colleagues highlighted how information encountered *early* can bias the evaluations of later information – leading to *primacy* biases in retrospective judgments (Bond, Carlson, Meloy, Russo, & Tanner, 2007; Russo, Carlson, & Meloy, 2006; Russo, Medvec, & Meloy, 1996; Russo, Meloy, & Wilks, 2000; Russo et al., 1998; see also Kardes, 1986; Kardes & Kalyanaram, 1992). When people express preferences for options (e.g., restaurants, dry cleaners etc.) whose attributes they learn via sequential displays, they tend to distort the evaluation of subsequent information on the basis of an emerging preference, which is established on the basis of the first few attributes. As soon as people develop a preference for option 1 over option 2 (because early attributes favoured option 1), they interpret any following information in a biased way in favour of option 1 (Russo et al., 1996; 1998).

In contrast to the above described "first takes all" perspective, recent findings have highlighted how the opposite pattern can also be true (Bruine de Bruin, 2005; see also Bruine de Bruine & Keren, 2003; Page & Page, in press). Bruine de Bruin (2005) performed a meta-analysis of the scores given by the judges of the Eurovision Song Contest and the European and World Figure Skating Championships. In both cases, the results suggested that the scores increased with the position of each performer; simply put, the later a candidate performed, the higher the chance of receiving high ratings. The general conclusion was that, once again, judgments provided when information is presented serially are biased by the order in which the decision-maker processes each piece of information. Performing at a later stage in the competition induced

more favourable ratings from the judges – a bias that can be generally labelled as a *recency* bias.

When people summarise information presented in a sequential manner, they seem to be biased by the *order* in which the information is provided to them. As for hedonic episodes, the relative importance of each piece of information (or moments within the episode) can be affected by *when* the information was encountered within the sequence (e.g., Anderson & Hubert, 1963; Bruine de Bruin, 2005; Hogarth & Einhorn, 1992; Russo et al., 1998). The most frequently observed biases are *primacy* and *recency* effects – that is, early or late information within a sequence is overweighed when summary assessments are provided.

Order effects in the social cognition literature

In the social cognition literature, order effects have been observed in evaluation tasks like personality impression formation (Anderson & Hubert, 1963; Asch, 1946; Dreben, Fiske, & Hastie, 1979; Lichtenstein & Srull, 1987), causality attribution (Collins & Shanks, 2002; Hogarth & Einhorn, 1992), and defendants' guilt (Pennington & Hastie, 1986; 1992; Reyes et al., 1980). *Early* (primacy) or *late* (recency) information has been found to disproportionately influence summary assessments.

For instance, in a study by Lichtenstein and Srull (1987), participants were presented with a sequence of behaviours. After the presentation, participants were asked to rate the likeability of the person who exhibited the behaviours. The extent of order effects was examined by relating these summary assessments to the number of positive (e.g. "lent money to a friend in financial

straits", p.102) and negative descriptions (e.g. "shot a songbird with his .22-calibre rifle", p.102) that were presented early and late in the sequence. Even if the total number of positive and negative statements was the same across trials, *primacy* bias was observed since likeability ratings were higher when positive information was presented early rather than late in the sequence. Primacy effects were also observed in the research run by Asch (1946; cited in Zauberman et al., 2006) and previously mentioned in section 1.2.1. The author asked participants about a person who was described as "envious", "stubborn", "critical", "impulsive", "industrious" and "intelligent". Other participants were presented with the same adjectives, but in the *opposite* order. The latter group of participants rated the hypothetical character more highly than the former group, as more positive adjectives were presented early in the sequence. Strikingly then, the same descriptive attributes could yield very different ratings depending on the order in which the descriptors were presented. When attributes with more positive connotations were given first, followed by less positive attributes, the person was rated more positively than when the order was reversed (Asch, 1946; cited in Zauberman et al., 2006).

On the other hand, in a study by Collins and Shanks (2002) participants were asked to perform a causal judgment task for which the evidence was presented in a sequential manner. Participants were mostly affected in their judgments by the most recent information (Experiment 1, p.1140), hence providing evidence for recency biases (see also Anderson, 1981; Hogarth & Einhorn, 1992).

The "Belief-Adjustment" model

In a review of these conflicting order effect results, Hogarth and Einhorn (1992) outlined the major factors that may play a role in determining whether primacy *or* recency effects will be observed in summary assessments. Their "Belief-Adjustment" model (1992) is based on a meta-analysis of the literature on summary assessments of sequential information; in this model, they draw a distinction between the *response mode* and the type of *processing* that a task involves. The response mode encompasses the type of judgment *task* people have to deal with. In *Step-by-Step* tasks, participants are asked to provide an evaluation after each piece of information is processed. On the other hand, in *End-of-Sequence* tasks participants are asked to express their evaluation only after all the relevant information has been presented to them.

At the same time, people can mentally *process* the information in two different ways. First, they can perform Step-by-Step processing, during which they review their opinion of the whole sequence after each piece of information is encountered. In Step-by-Step processing, the judgment is *adjusted* as many times as there are steps in a sequence of information, something that is thought to require a relatively high cognitive effort. Most importantly, according to the Belief-Adjustment model, Step-by-Step processing will lead to *recency* effects in summary assessments, as the last information will have approximately the same weight in the overall evaluation as *all* the aggregated information that preceded it. Conversely, in End-of-Sequence 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). According to the authors, End-of-Sequence processing will

therefore lead to *primacy* effects, as early information has roughly the same impact on summary assessments as *all* the information that followed it.

Importantly, the way the participants mentally *process* information (processing) does not necessarily have to reflect the task instructions (response mode). If the task is Step-by-Step, participants have to perform Step-by-Step processing – as they are asked to provide a judgment after each piece of information. On the other hand, if the task is End-of-Sequence, then participants are free to rely on either Step-by-Step or End-of-Sequence processing. The Belief-Adjustment model suggests that this choice is determined by other task-related variables like the *familiarity* of the task, and the *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 Step-by-Step processing; hence recency effects should be observed. On the other hand, simple and short information (or familiar tasks) would be associated with End-of-Sequence processing – and consequently with primacy biases.

With respect to the present thesis, an important consideration should be taken from the above review. In the Belief-Adjustment model (Hogarth & Einhorn, 1992) there is no explicit mention or description of the role that memory processes may play in biasing retrospective evaluations – a question which is at the core of the work reported here.

1.3. The role of memory

1.3.1. Parallelisms between memory and judgment bias

Why are people biased when evaluating the past?

In light of the phenomenology described above, a question that naturally arises is the following: *Why* do biases in summary assessments arise? Despite the divergence of the reviewed findings, a general principle can be extrapolated: People overweigh some moments of an experience – or information within a sequence – when assessing it in hindsight. *Why* is that? It can be argued that psychological models of these biases should provide a causal explanation for these phenomena in terms of the cognitive processes that generate them.

The literature on the "Peak-End" rule (e.g., Fredrickson & Kahneman, 1993) does not *empirically* address the 'why' of biases in retrospective evaluations. "Peak" and "End" are simply defined as "gestalt" features of an event; a full explanatory interpretation about their effects is clearly missing (Ariely & Carmon, 2000; Loewenstein & Prelec, 1993; Read, 2004). Fredrickson (2000) offers the following post-hoc discussion of Peaks and Ends "…earn their privileged status because they carry more personal meaning than other moments" (p. 589). In the same paper, the author claims that the affect experienced during the Peak moment "…is the single moment that defines the personal capacity needed to face the experience again" (p. 590). Although interesting, these attempts to explain the impact of Peaks and Ends rely on general post-hoc ideas, which remain untested. Similarly, Ariely and Carmon (2000) put forward the notion of *extrapolation*, according to which people can predict how experiences

will unfold in the future by summarising specific features of the experience at hand (e.g., trend and end). Also, extracting just a few moments of the experience while summarising it in hindsight might benefit people as this process may counter their limits in *encoding* capacity. Once again, though, no detailed explanatory account of the biases has been suggested (but see Hogarth & Einhorn, 1992).

In the present thesis, I examine an account of biases in retrospective evaluation which relies on well-known cognitive phenomena, i.e. *memory* operating characteristics.

Judgment bias as a memory bias

When reviewing the biases in retrospective evaluations, striking parallels can be drawn with well documented memory phenomena. For instance, the effect of a "Peak" on retrospective evaluations can be associated to its *distinctiveness* relative to the other moments in the to-be-assessed experience. In the memory literature, distinctiveness is a relational feature in that it relates to "the extent to which an item stands out from or is distinct from the other items in the search set" (Neath & Surprenant, 2003; p. 458). In a to-be-recalled list, items which are distinctive are usually better recalled than the surrounding items (e.g., Hulme, Neath, Stuart, Shostak, Surprenant, & Brown, 2006; Hunt, 1995; Hunt & Lamb, 2001; Nairne, 2005; Neath, Brown, McCormack, Chater, & Freeman, 2006). Consider the following experiment on the isolation effect (or "Von Restorff" effect) (Rabinowitz & Andrews, 1973). Participants were asked to recall 11-item lists; the items were presented on cards and they were typed in black capital letters. In some lists, one of the words was presented in larger capital red letters,

which undoubtedly made it perceptually distinctive compared to the remaining ten items. The results were clear-cut: When isolated, the recall of the distinctive item was dramatically increased compared to its neighbours – with the recall advantage over non-isolated items presented in the same position reaching approximately 35%.

A simple memory-based approach could then explain the influence of "Peaks" on retrospective evaluations. The logic is as follows: "Peaks", or the most intense moments of an experience, are by definition more distinctive compared to the other (less intense) moments. As "Peaks" are distinctive, they are more likely to be retrievable from memory at the time a summary evaluation is required. If memory plays a central role in biasing retrospective evaluations, then this recall advantage for the "Peaks" should translate into a larger impact on summary assessments – compared to the remainder of the experience.

Moreover, if we were adopt a memory-based approach, primacy and recency effects observed for summary assessments (see Hogarth & Einhorn, 1992) can be related to primacy and recency effects in memory, i.e. the recall advantage for early and late information within a series (e.g., Murdock, 1962; Neath & Surprenant, 2003; Tan & Ward, 2000). Just to re-iterate the above described logic, in a number of circumstances, we could expect the first and last moments of an experience (or information within a sequence) to enjoy a recall advantage compared to intermediate moments – which could explain their larger impact on summary assessments.

Situational or task-related factors may moderate the extent to which either primacy *or* recency information is more prominent in memory. In turn, this could explain why either early *or* late information has been observed to influence

summary assessments (e.g., Anderson, 1981). For instance, if the delay between the *study* phase (the time during which some to-be-recalled items are learnt) and the *test* phase (when participants recall the items) is increased, the recall for early items increases while the recall for late items decreases (the "recency–primacy shift"; e.g., Knoedler, Hellwig, & Neath, 1999). Also, the modality of presentation of the to-be-recalled (and to-be-evaluated) information may influence its memorability: Late information is usually better recalled if a sequence is presented aurally rather than in the visual modality (e.g., Corballis, 1966; Murray, 1966).

To summarise, by referring to the influence of memory processes, a comprehensive account of judgment bias in retrospective evaluations can potentially be developed. Moreover, such a theoretical approach would go some way towards bridging together the research on memory on the one hand and on judgment on the other. A memory-based approach can probably not provide a complete account of all the phenomena reported in retrospective assessments; however, it appears clear from the above and from the work reported hereafter that considering memory operation and retrieval dynamics is likely to contribute to a better understanding of judgment in hindsight. Moreover, previous research has neglected the role of memory in decision making, prompting Dougherty, Gronlund and Gettys (2003) to claim that "research on behavioural decision theory has largely failed to address the memory processes underlying judgment and decision making" (p. 125; see also Weber, Goldstein, & Barlas, 1995; Weber, Goldstein, & Busemeyer, 1991; Weber & Johnson, 2006).

1.3.2. Availability and accessibility

Evidence of memory-driven judgment bias

Although there has been very little systematic investigation into the role of mnemonic processes in retrospective judgment of hedonic episodes, there is no paucity of hypotheses that suggest this would be a worthwhile endeavour (e.g., Higgins, Rholes, & Jones, 1977; Loken & Hoverstad, 1985). Taylor (1982) claims that judgment *must* depend on memory processes, as "one's judgments are always based on what comes to mind" (p. 199). Indeed, several theoretical models maintain that memory processes, at least to some extent, influence people's decision-making (e.g., Fuzzy-trace theory; Brainerd & Reyna, 1990; MINERVA-DM; Dougherty, Gettys, & Ogden, 1999; Decision by Sampling; Stewart, Chater, & Brown, 2006; see also Kashima & Kerekes, 1994).

What is more, several empirical findings support the suggestion of a close association between memory and judgment processes. In their seminal work on the *availability* heuristic, Tversky and Kahneman (1973) asked participants to judge whether some letters (e.g., "k") appear more often either in first or third position within English words. Despite the fact that in English there are more words with a "k" in third position than in first position, the overwhelming majority of participants reckoned the opposite was true. The authors argued that these biased frequency estimates were due to a memory-based heuristic; namely, participants brought to mind – i.e., they retrieved from memory – words that started with a "k" (e.g., "kitchen") *more easily* than words with a "k" in third position (e.g., "acknowledge").

Similarly, people tend to overestimate the occurrence of violent deaths, presumably because they are more thoroughly covered by media reports and

hence more available in memory (Combs & Slovic, 1979; Lichtenstein, Slovic, Fischhoff, Layman, & Combs, 1978). The work by Dougherty and colleagues (Dougherty, 2001; Dougherty & Franco-Watkins, 2003; Dougherty & Hunter, 2003a; 2003b; Dougherty & Sprenger, 2006; Sprenger & Dougherty, 2006) has suggested that working memory is an important determinant of hypothesis generation and probability judgment. For example, in a series of studies, it was showed that participants with higher working memory capacity managed to generate more alternative hypotheses – and consequently performed probability judgments more accurately (Dougherty & Hunter, 2003a).

Finally, a study by Reyes et al. (1980) investigated whether changes in the memory trace due to delay may result in different judgments about the same information – depending on when the evaluation is prompted. In their study, participants partook in a mock trial during which a defendant was charged with drink driving. Participants were presented with arguments, half of which were in favour of the prosecution, while the remaining half were in favour of the defendant. Each of these arguments was presented either in a vivid or a non-vivid form. For instance, an argument that described how the defendant was drunk before leaving the party was non-vividly formulated as follows: "On his way out the door, Sanders [the defendant] staggered against a serving table, knocking a bowl to the floor." (p. 4). The vivid version was instead: "On his way out the door, Sanders staggered against a serving table, knocking a bowl of guacamole on the white shag carpet." (p. 4). There were two groups of participants; the first group heard the prosecution arguments in vivid form and the pro-defendant arguments as non-vivid. The second group heard the opposite. As predicted by the authors, there was no difference between the two groups in their judgment of

the defendant's guilt – when participants were tested *immediately* after the arguments presentation. This was attributed to participants' accurate memory of all the arguments, which, to reiterate, were half in favour of the defendant and half in favour of the prosecution. However, when participants were tested *two days later*, participants who saw the prosecution arguments in vivid form judged the defendant more severely than the other participants. The authors argued that, as some time had passed, the vivid arguments were "more available for recall and hence they should dominate the non-vivid counterarguments." (p. 3). The results of a memory test supported this prediction, as vivid arguments were better remembered than non-vivid arguments. Also, across participants there was a correlation between the judgment about the defendant's guilt and the recall advantage for prosecution arguments – as compared to pro-defendant arguments.

Retrieval content vs. Ease of access

The above reviewed studies also introduce an important debate about *how* memory processes may influence judgment. The work by Schwarz and his colleagues (Schwarz, 1995; 1998; Schwarz, Bless, Wanke, & Winkielman, 2003; Schwarz & Vaughn, 2002; see also Tybout, Sternthal, Malaviya, Bakamitsos, & Park, 2005) has highlighted that memory can affect judgment because of its *retrieval content* or because of its *ease of access* ("retrieval fluency"; cf. Benjamin & Bjork, 1996; see also the Accessibility model; Koriat, 1993; 1995; Koriat & Levy-Sadot, 2001)[1]. The former (retrieval content) refers to the fact

[1] Schwarz and colleagues (e.g., Schwarz, 1998; Schwarz & Vaughn, 2002) experimentally investigated how the effects of retrieval content can be disentangled from those of ease of access. Normally, people seem biased by the ease or difficulty with which information is brought to mind. However, people rely more on *what* they can retrieve from memory "when the subjective recall experiences are considered non-diagnostic" (Schwarz & Vaughn, 2002; p. 118).

that people base their judgments on *what* comes to mind (e.g., Higgins, 1989; 1996). For instance, in the above study by Reyes et al. (1980), participants were more inclined to judge the defendant as guilty when they remembered *more* arguments in favour of the prosecution. Bias in judgment would then stem from people's tendency not to retrieve all the presented information in order to evaluate it; rather, they seem to provide a judgment on the basis of the sub-set of information that comes to mind most readily (e.g., Bodenhausen & Wyer, 1987; Higgins, 1996).

On the other hand, the formulation of the availability heuristic (Tversky & Kahneman, 1973) states that it is the *experienced* ease of retrieval that determines judgment. In the example mentioned above, participants overestimated the frequency of words with a specific letter in first position as it was *easier* for them to retrieve such exemplars from long-term memory – as compared to items where the letter appears in third position (Tversky & Kahneman, 1973; see also Lichtenstein et al., 1978).

So far, the evidence and theoretical accounts reviewed support the general claim that memory processes influence judgment. However, several findings from the social cognition literature (e.g., Hastie & Park, 1986; but see Betsch et al., 2001) question the value and generality of a memory-based approach.

1.3.3. Independence accounts

Evidence against the role of memory in evaluations

If the evaluation of the past is based upon our memory of events it follows that memory and judgment must be functionally related at some level. Therefore, if people are asked to evaluate an experience and they are also tested for their memory of it, clear associations should be observed between judgment and memory measures. That is not always the case. In their influential study on personality impression formation, Anderson and Hubert (1963) instructed participants to rate the likeability of a hypothetical character who was described through a series of adjectives. The authors also asked participants to perform a memory task, which involved *recalling* the adjectives. The results did not display the expected association between memory and judgment measures. On one hand, the findings revealed a primacy bias in that early adjectives in the list had more influence on the likeability ratings. On the other hand, a recall advantage was displayed for the last adjectives (*recency* effect). The results suggested a dissociation between memory and judgment processes, in that the most memorable adjectives were not those which biased the ratings. The authors formulated what later on was going to be defined as the "two-memory hypothesis" (cf. Hastie & Park, 1986) by concluding that "the impression response is based on a different memory system than that which underlies the verbal recall." (p. 388).

Other empirical evidence also points towards the functional independence of memory and judgment (e.g., Banaji, Hardin, & Rothman, 1993; Dreben et al., 1979; Fiske, Taylor, Etcoff, & Laufer, 1979). For example, the *exposure effect*, which refers to the increased preference for some items only because they are

repeatedly presented, can occur in the absence of increased *recognition* for those specific items (Zajonc, 1980).

More recently, Betsch et al. (2001) proposed the "Value-Account" model, which explains these dissociations between memory and judgment with a logic similar to that of the "two-memory hypothesis" (Anderson, 1981; 1989; Anderson & Hubert, 1963). In their study, participants were presented with an *incidental* judgment task. They were told that the experiment was about their memory for pictorial ads presented on the screen, one at the time. They were also told to read the return values of five different stock exchange shares, which were displayed together with the ads. At the end of the presentation they were asked to both estimate the sum and average return of each share and to express their attitude towards them, answering the generic question "How do you find this share?" (p. 245) Despite being rather inaccurate in estimating the return of the shares, participants' attitudes were most favourable towards the most profitable shares. The authors interpreted these findings as evidence of a dissociation between memory and judgment. They suggested that participants shaped their attitude towards the shares *while they were presented with the relevant information*; at the end of the presentation, their impression of each share was stored in a hypothetical structure (the "value-account"). When prompted with a judgment task, the value-account was more easily accessible than the *actual* information about the shares; this in turn led to the dissociation between participants' estimations of the shares and their attitudes towards them (Betsch et al., 2001).

On-line vs. memory-based judgment

The literature reviewed in the previous sections provides us with conflicting views about the role of memory in summary assessments. Does judgment depend on memory processes?

In their seminal work about the relationships between memory and judgment, Hastie and Park (1986) provide an interpretation of these inconsistent findings. The authors suggest that, when providing summary assessments, people can engage in one of two information processing strategies. If a *memory-based* strategy is used, relevant information is retrieved from memory and integrated to produce an evaluation. When this is done, correlations between memory and judgment are observed, as the latter depends on retrieval processes (e.g., Dougherty et al., 1999; Schwarz, 1998; Tversky & Kahneman, 1973). Conversely, if *on-line* processing is relied upon, an impression of the target stimulus is formed *while* it is being encountered; this impression is adjusted after each piece of new evidence is processed. When an overall evaluation is elicited, this impression formed on-line is retrieved and it forms the basis of retrospective evaluations. Correlations between memory and judgment are therefore not necessarily observed, as the summary assessment does not rely on specific information about the target stimulus retrieved from memory (e.g., Anderson, 1996; Betsch et al., 2001; Lichtenstein & Srull, 1987). On-line judgment seems to refer to a cognitive process which is rather similar to the Step-by-Step processing included in the Belief-Adjustment model (Hogarth & Einhorn, 1992).

According to Hastie and Park (1986; see also Hastie & Pennington, 1989), on-line judgment is the predominant form of processing people implement when evaluating reality. In support of this claim, experimental evidence has

suggested that trait (e.g., Winter & Uleman, 1984) and causal inferences are performed spontaneously and on-line (Weiner, 1985). Moreover, it is argued that, as the on-line judgment is readily available, searching memory for specific information about the to-be-assessed target stimulus is a cognitively costly and unnecessary process. Hastie and Park suggest that memory-based judgments are relatively rare; they are performed when there is no forewarning about the judgment task, and when people are not able to spontaneously form an impression. In these cases, as the on-line judgment is hindered, people *have to* rely on the information they stored in memory about the to-be-assessed target in order to evaluate it. In one of their studies, Hastie and Park (1986) had participants listen to taped sentences which described actions performed by a person. Some of the actions were indicators of the person's *intelligence*, *friendliness* and *likeability*. In order to prevent spontaneous personality impression formation, the authors asked one group of participants to rate the correctness of the grammar of each sentence before evaluating the person's intelligence, friendliness and likeability (memory-based judgment condition). The other group of participants were instead told to simply attend to each of the sentences (on-line judgment condition). As predicted by the authors, correlations between memory for the sentences and the evaluations of the character's personality were observed only for the memory-based group (see also Bargh & Thein, 1985; Beauvois & Cambon, 1998; Bizer, Tormala, Rucker, & Petty, 2006; Custer & Hurts, 2003).

To sum up, while several findings suggest that memory and judgment may be functionally related (e.g., Tversky & Kahneman, 1973), other evidence points towards a different conclusion and supports "Independence" accounts

(e.g., Anderson, 1981). In the next sections, we will argue that a number of theoretical and methodological issues must be addressed before the memory–judgment relationships can be adequately understood.

1.4. Summary and overview of the thesis

1.4.1. Summary

General objectives

The present thesis aims to systematically investigate whether memory processes may be underlying some of the above described biases in retrospective evaluations (e.g., "Peak" and "End" effects; Fredrickson & Kahneman, 1993; Primacy and recency effects; Anderson & Hubert, 1963; Collins & Shanks, 2002). In doing so, we are attempting to bridge together different literatures which reported biases in summary assessments – namely the behavioural decision making, the social cognition and the information integration literatures (e.g., Ariely & Carmon, 2000; Hogarth & Einhorn, 1992).

Issues that need to be addressed

In order to thoroughly assess whether memory processes are related to evaluations, a series of methodological and theoretical issues must be addressed. Although an overview of these issues is reported and discussed in each of the following experimental chapters, some of the more important questions that motivated the present work are summarised here.

One important issue relates specifically to the biases in retrospective evaluations. The literature on the Peak-End rule, when predicting retrospective evaluations, often did not isolate early moments (primacy) as one of the potential predictors (e.g., Ariely & Carmon, 2000). At the same time, the paradigms used in the social cognition literature did not allow one to disentangle the effects of "Peak" information from those of the order of presentation (e.g., Hogarth & Einhorn, 1992). In the studies reported here, it was possible to disentangle and investigate the separate contributions of primacy, recency and "Peak" information on judgments.

When investigating the relationships between memory and judgment correlations, previous research mostly adopted a simple correlational approach (Hastie & Park, 1986). Memory and judgment measures were collected, and any possible association was investigated *post-hoc*. In the present thesis, *a priori* hypotheses were put forward about judgment outputs depending on the memorability of specific information.

Moreover, research that attempted to directly assess the *causal* nature of the relationship between memory and judgment mainly investigated *frequency* judgments – during which people judge the frequency of a target stimulus or the likelihood of its occurrence (e.g., Gabrielcick & Fazio, 1984; Hanita, Gavanski, & Fazio, 1997; Lewandowsky & Smith, 1983). In the present thesis, we will investigate instead *quality* judgments, where people summarise the hedonic value of affective stimuli. This is an important distinction as it is likely that different cognitive processing is underlying the two judgment tasks – namely frequency and hedonic judgments. As Hogarth and Einhorn (1992) suggest in their "Belief-Adjustment" model, frequency judgments fall into the category of *evaluation*

tasks (p. 9). That is, people have an *a priori* hypothesis (e.g., a given theory is true), and any incoming information is encoded as either supporting or not supporting such hypothesis. The data for this type of judgment task can be accounted for by additive models, where the final judgment (bipolar itself) depends on the total amount of information in favour of the hypothesis as compared to the amount of the information that disproves the hypothesis. On the other hand, in *estimation* tasks (p. 9) people provide impressions which reflect the *degree* to which something (or someone; Anderson & Hubert, 1963) is liked. In these cases, which more closely reflect summary assessments for hedonic events, averaging models seem to be more appropriate in order to predict people's judgment (Hogarth & Einhorn, 1992).

Compared to previous research on the relationships between memory and judgment (e.g., Hastie & Park, 1986), more comprehensive and detailed memory measures are used here. It is argued that these are necessary if the influence of memory processes on judgment is to be detected. More specifically, memory measures relating to *availability* were used; availability refers to *what* is accessed in memory at the time of judgment (e.g., Schwarz, 1998). We also used additional measures in order to detect any potential effect of *accessibility* – namely the *ease of access* of information in memory (e.g., Tversky & Kahneman, 1973).

A further methodological improvement involves controlling for variables that may be affecting retrieval of information from memory. For instance, stimulus frequency and familiarity may influence retrieval of verbal information; at the same time, though, these factors may not affect the processes underlying retrospective judgments (e.g., likeability of a person). This, in turn, could reduce

the possibility of detecting any underlying correlation between memory and judgment outputs. Thus, by controlling variables like stimulus frequency and familiarity, it is argued that any conclusions about the relationships between memory and judgment are more reliable.

1.4.2. Thesis outline

Three groups of studies are reported, each of them addressing different but related theoretical questions. Chapter 2 investigated the judgment bias in retrospective evaluations of short and discrete affective episodes, namely lists of words. More importantly, it addressed the open question whether a reliable relationship between memory and judgment can be observed. Compared to previous research, some important methodological improvements were implemented.

As a clear association between memory and judgment was found, Chapter 3 investigated further the nature of the relationship. Biases in retrospective evaluations were *predicted* on the basis of the memorability of the to-be-assessed information, which was experimentally manipulated. Also, this study addressed the potential confounding effects of the dual nature of the paradigm, where participants are asked to both evaluate and recall the presented information.

Chapter 4 extended the above findings to less abstract and more cohesive experiences, i.e. short stories told through slideshows. The outlined memory-based approach also held for this type of stimuli and accounted for the biases in retrospective evaluations – which differed from those observed in the previous two chapters. Finally, this series of studies provided some support for the idea that the choice of the memory task is important when studying the relationship

between memory and judgment. More specifically, it is suggested that if the memory task does not call for the retrieval of the same information called upon for judgment, the relationship between judgment and memory can be obviated. The argument is basically that previous null effects could be attributable to misalignment between the judgment and memory tasks in terms of what is retrieved to comply with the demands in each task.

Finally, in Chapter 5, the findings reported here are summarised and discussed in light of recent memory-driven theories of decision making. Also, future directions for the research on the role of memory in judgment and decisions making are suggested.

Chapter 2:
The Role of Memory in Retrospective Evaluations

Introductory note

In the following pages, three experiments are reported, in journal article form. This paper represents the first empirical chapter of the thesis – and it has been submitted for publication. The objectives of the present experiments are two-fold. First, the nature of the biases in retrospective evaluations for short episodes is explored. Second, it is investigated whether memory and judgment measures are associated.

2.1. Abstract

Current research shows that when evaluating experiences retrospectively, people tend to base their assessments on specific moments of the experience. For instance, the Peak-End rule (e.g., Fredrickson & Kahneman, 1993) states that the most intense ("Peak") and final moments ("End") of an event exert a large influence on retrospective evaluations, while more holistic features (e.g., event duration) have little impact. In the present series of studies, the role memory plays in biasing retrospective evaluation was explored. Participants recalled word lists after having assessed each of them for pleasantness. The results showed that the assessments were significantly influenced by the presence and the position of a negative item. Primacy and recency effects were observed for both memory and judgment measures. Further analyses confirmed that recalling the negative information was associated with more unpleasant judgments. The relationship between memory and judgment was clearest in the last experiment, where the memory task demands were lowered. The results are discussed in relation to a memory-based approach to the judgment biases in retrospective evaluations.

2.2. Introduction

It is a common experience to reflect on past events, evaluating them in hindsight. When such evaluations are reported, people tend to provide a summary assessment of the episode under consideration instead of a moment-by-moment account of how they felt during the experience (Fredrickson, 2000). The event may have been long-lasting and changed in quality over time (e.g., a lecture, a holiday etc); nevertheless, evaluations typically involve a cohesive unitary response. How do people summarise the quality of an experience as a whole into a global assessment? That is, how do they construct retrospective evaluations?

Recent findings have highlighted the importance of better understanding the processes that underlie retrospective evaluations. For example, Kahneman (2000) showed that an event's *remembered* utility—i.e. how it is assessed in hindsight—better predicted decisions to repeat the experience than its *actual* experienced quality or experienced utility (measured while the event was taking place). Kahneman's view highlights the idea that retrieval from memory may play a determining role in the construction of retrospective evaluations. The aim of the work reported herein was to appraise this memory-based view.

Although it might seem obvious on first examination that memory must play a significant role in retrospective evaluations many proposals do not adopt this premise. In effect, there is considerable controversy in the literature regarding the role that memory plays in retrospective evaluations. For instance, "on-line" theories hold that people evaluate events as they experience them (e.g., Anderson, 1981). According to this view, to judge an event, the only thing needed from memory is the result of the on-line assessment. Retrospective

evaluations are thought to be based on this on-line processing; retrieving information about the event itself is seen as cognitively costly and unnecessary. This approach is supported by a number of studies that found no relationship between what is remembered from an episode and its retrospective evaluation (e.g., Betsch, Plessner, Schwieren, & Gütig, 2001); we will return to these findings below.

On the other hand, theoretical approaches in line with the "Availability Heuristic" (Tversky & Kahneman, 1973) predict that retrospective evaluations will depend on the information that can be remembered about the event – and the ease with which it can be accessed in memory. Results which highlight correlations between memory and judgment support this view (e.g., Reyes, Thompson, & Bower, 1980).

In the present paper, we set out to contribute to this debate by systematically investigating the role of memory in retrospective evaluation. Relative to previous work, our approach has a number of novel characteristics. Our strategy was to design an experimental paradigm where we could measure a) retrospective evaluations and b) memory for the assessed material. Importantly, we adopted a transfer-appropriate-processing perspective: it is suggested that conflicting results were obtained in the past at least in part because memory is surprisingly flexible; what is remembered is highly dependent on the specific retrieval cues provided at the point of retrieval. This fact is easy to underestimate—however, one has only to consider the findings on false memory or the misinformation effect to realise that small changes in the cueing information can produce big changes in what is remembered (see Loftus, 2005, for a review). If retrospective evaluation is obtained in one context and memory

for the material is assessed within a different procedure this may well alter the cueing environment enough to significantly reduce any relationship between memory and judgment. Hence, if one is to clarify the role that memory plays in retrospective evaluation, it is important to show that how memory is probed is aligned with the information that underlies the judgment task. This objective was pursued here and was the dominant motivation for our choice of task and materials, as will be explained below.

Another important element of our strategy involved manipulating the memorability of the material. Based on a perspective where memory plays an important role in judgment in hindsight, we predicted the effect memorability would have on retrospective assessment. Therefore, contrary to several previous studies on the relationships between memory and judgment (for a review, see Hastie & Park, 1986), the causal role of memory processes could be more confidently assessed.

Finally, in the studies reported below, we systematically considered *accessibility*, i.e. the ease with which information is accessed in memory (see also Schwarz, 1995). This allowed us to test the hypothesis that accessibility moderates the impact of memory on retrospective evaluations; if this is the case then the easier it is to retrieve a given feature or part of an experience, the larger its effect on retrospective judgment.

Before providing more detail relative to the above, we first briefly review the relevant literature on retrospective evaluation as well as the work that relates to the role of memory in the process.

Judgment biases in retrospective evaluations (RE)

A number of studies have investigated how people construct or develop retrospective evaluations (cf. Fredrickson, 2000). This research examined retrospective evaluations in many domains. The latter have included personality impression formation (Anderson & Hubert, 1963; Asch, 1946; Dreben, Fiske, & Hastie, 1979; Lichtenstein & Srull, 1987), perception of pain (Ariely, 1998; Ariely & Carmon, 2000; Redelmeier & Kahneman, 1996), the profit profile of market shares (Betsch et al., 2001), payment sequences (Langer, Sarin, & Weber, 2005), movie clips (Fredrickson & Kahneman, 1993), video output quality (Hands & Avon, 2001), causality attribution (Hogarth & Einhorn, 1992), perceived guilt of defendants (Reyes et al., 1980), aversive experiences (Schreiber & Kahneman, 2000; Varey & Kahneman, 1992) and TV advertisements (Baumgartner, Sujan, & Padgett, 1997).

One of the conclusions of this work has invariably been that judgment in hindsight it biased in various ways. For example, in a number of social judgment tasks, order effects have been observed (e.g., Collins & Shanks, 2002; Pennington & Hastie, 1986; for a review, see Hogarth & Einhorn, 1992). Early (primacy) or late (recency) information has been found to disproportionately influence summary assessments. Consider a well-known study by Anderson and Hubert (1963). In this study, participants were presented with a sequence of adjectives. Participants were then asked to rate the likeability of the hypothetical person whom the adjectives referred to. Even if the number of positive and negative statements was held constant across sequences, primacy effects were observed: likeability ratings were higher for those lists where the positive information was presented early in the sequence and significantly lower when the

negative descriptors were presented first. These order effects suggest that certain moments of the assessed episode weigh more heavily in retrospective evaluations than others, biasing the assessments provided. This is also a conclusion that can be drawn from another important body of work—the one examining what is known as the Peak-End rule.

The "Peak-End" rule (Fredrickson & Kahneman, 1993) proposes that the final moments and the most intense[2] parts of an experience disproportionately affect judgments in hindsight (Ariely & Carmon, 2000; Baumgartner et al., 1997). The Peak-End rule's predictions have been supported with many types of stimuli: Positive (Do, Rupert, & Wolford, 2008) and negative experiences (Kahneman et al., 1993), very short (Schreiber & Kahneman, 2000) and prolonged events (Clark & Georgellis, 2004; Redelmeier et al., 2003), continuous (Baumgartner et al., 1997; Kahneman et al., 1993) and discrete episodes (Clark & Georgellis, 2004; Langer et al., 2005), and finally experiential (Ariely, 1998) and semantic content (Fredrickson & Kahneman, 1993). Although some issues remain controversial, the work on the Peak-End rule clearly supports the general point that judgment in hindsight is biased. Moreover, we would suggest that in many instances, these biases appear connected to well known characteristics of memory functioning, a suggestion to which we now turn.

What is the role of memory? Availability Vs On-line judgment

One of the important questions motivating the present investigation was: *why* do certain parts of what is assessed exert a larger impact on retrospective evaluations than others? When consulting the literature on memory

[2] Intensity is defined as the magnitude of the hedonic dimension that is being evaluated (e.g. pain, pleasantness etc.)

phenomenology, striking parallelisms with the above mentioned judgment biases are easily identifiable. For instance, distinctiveness (i.e. the extent to which a piece of information stands out against the background created by associated material; Neath & Surprenant, 2003) usually leads to a memory advantage in standard memory tasks (e.g., Hunt, 1995). If we were to adopt an approach to retrospective evaluations biases where memory plays an important role, then it could be suggested that the most intense moments of a just-experienced event are more likely to be distinctive – relative to the background of less intense moments comprising a given experience. It is an easy step to then suggest that these more intense moments are also going to be most readily available in memory. Consequently, it seems reasonable to argue that such a memory advantage could result in a larger impact for the distinctive moments on summary assessments. According to the same line of reasoning, primacy and recency effects in memory (e.g., Tan & Ward, 2000) would be reflected in primacy and recency effects in retrospective judgment.

An implication of the memory-based approach, in line with the availability heuristic (Tversky & Kahneman, 1973), is the prediction of a relationship between memory and judgment: The information that is most readily accessible in memory at the time of judgment should weigh more heavily in summary global assessments. Several empirical findings (e.g., Higgins, 1996) have pointed in that direction and revealed how people perform their judgment depending on the episodic information that is most readily available in memory, without retrieving all the information they could about a just-experienced event or information sequence.

However, a number of studies from the social cognition literature have *not* found any correlations between memory and judgment measures (for a review, see Hastie & Park, 1986). For example, in the above mentioned study by Anderson and Hubert (1963), participants provided likeability ratings of hypothetical characters after being presented with a series of adjectives. Having asked participants in some conditions to *recall* the adjectives as well, the authors could relate judgment and memory measures for the same information. Strikingly, while a strong primacy effect was observed for the likeability ratings, recency effects were noted for the recall of the adjectives. In this case, what seemed most memorable (the most recently presented adjectives) was not related to what was biasing evaluations (the positive or negative information that came first on the list of adjectives).

Some theoretical approaches have attempted to account for these dissociations by proposing functional independence between two cognitive processes (Hastie & Park, 1986). The "two-memory hypothesis" (Anderson, 1981; Anderson & Hubert, 1963) suggests that evaluations of events or sequences and the storage of their verbal traces into long-term memory are two independent processes, that are differentially called upon depending on the type of task participants are asked to perform. Simply put the suggestion is that the evaluation and integration of information necessary for the judgment task is performed in parallel to the encoding of the event information as such into long-term memory. Because both sources of information are independently encoded, they are also accessed separately, producing the dissociations referred to above. A similar approach, also hypothesising independence between memory and judgment processes, is the "value-account" (Betsch et al., 2001). Betsch and

colleagues hold that the summary evaluation of any episode is formed on-line, i.e. while the event is being experienced. At the end of the event, this impression is stored in a hypothetical structure (the "value-account") which is more easily accessed when a judgment is required than the memory trace of the event itself.

In order to explain the contradictory findings relating to memory-judgment correlations, Hastie and Park (1986) insisted on an analysis of the judgment task itself. According to these authors, in almost all the studied situations participants are either aware of the subsequent judgment task or are capable of performing spontaneous judgments; therefore, it is possible to form an impression of the event while it is being experienced (on-line judgment). The implication is that there is no need to retrieve information from memory about the event itself to provide a retrospective evaluation, since retrospective judgment can be based on the impression formed on-line. It follows that the only situations in which participants would rely on episodic information about an event in order to assess it retrospectively, are the instances where they are neither able to form a spontaneous judgment nor forewarned of a subsequent judgment task. Arguably, this would exclude many if not most of the experiences we encounter in our everyday lives.

Here, we adopt a position that is not so removed from the basic perspective adopted by Hastie and Park (1986): The information retrieved in a typical verbally based recall task can be significantly different from the information that a retrospective evaluation was based on. However, we do not think this necessarily leads to the conclusion that retrospective judgment is independent of episodic memory. Rather, we would suggest that this is an argument for carefully considering how the judgment and memory tasks are

aligned. From the perspective of experimental memory research, this perhaps has not been given all the needed consideration in the past. We address this issue in the experiments presented here.

Moreover, from a more methodological perspective, it is argued that a number of issues must be addressed before the memory-judgment relationship can be adequately understood. For example, together with several other studies which investigated memory-judgment correlations (e.g., Lichtenstein & Srull, 1987; Reyes et al., 1980) Hastie and Park (1986) used a ratio measure for the memory output. For instance, 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. The authors then computed a memory ratio 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 possibly each item on participants' mind at the time of judgment) has the same weight in the overall evaluation. Such an assumption is disputable in light of the work reviewed above showing significant biases in the weight various elements have in retrospective judgments. In addition, the availability heuristic (Schwarz & Vaughn, 2002; Tversky & Kahneman, 1973) suggests that the ease of recall of a given piece of information moderates its effect on evaluations. Accordingly, it can be argued that the availability of each item at the time of judgment will impact on retrospective evaluations: the more

easily accessible a specific item is at the time of evaluation, the higher its influence on the evaluation itself. This would in turn entail that a memory measure (like a ratio) which assigns the same weight to each item recalled does not adequately represent the memory content accessed in order to produce overall assessments. Finally, by using a positive divided by positive + negative ratio as the memory measure, the memory output for a participant who recalled 4 positive and 2 negative items is identical to that of a participant who retrieved 2 positive and 1 negative item. Hence, the information about the accuracy of the memory representation at the time of its evaluation is ignored.

A second methodological issue is worthy of mention. To our knowledge none of the previous studies in the social cognition literature which investigated the correlations between memory and judgment measures controlled for variables which are known to affect memory performance (e.g., stimulus frequency and familiarity). It seems reasonable to suggest that such factors could affect the retrieval of verbal information while not influencing the processes that underlie retrospective evaluations of the same information sequence. This could in turn hinder – at least in some studies – the underlying correlation between memory for a specific event and its retrospective evaluation. For instance, in their study on personality impression formation, Dreben et al. (1979) utilised sentences relative to certain behaviours to describe hypothetical characters. Even if control over the valence of stimuli (which is thought to affect the judgment task) was exercised, the authors do not mention control over length, familiarity or frequency of target items - factors all well known to influence memory processes. Likewise, Anderson and Hubert (1963) used sequences of trait adjectives: Once again stimuli were selected depending on their diagnostic

quality (e.g., being “favourable”, “medium” or “unfavourable”, p. 380), and none of the selection criteria involved parameters which could well be affecting memory performance. In the experiments described below, we addressed these design issues.

Overview of Experiments

The main objective of the present paper was to explore the predictions of a memory-based approach to judgment bias in retrospective evaluations. In doing so, we investigated the relationships between memory and judgment and tested the predictions of two contrasting approaches. On the one hand, theoretical accounts that suggest independent processes underlying memory and evaluation processes (Anderson, 1981; Betsch et al., 2001; Hastie & Park, 1986) predict experimental dissociations between memory and retrospective evaluations. On the other, recent work as well as more classic papers has highlighted the biases that are present in summary assessments – biases that indicate that certain moments or characteristics of an event weigh much more heavily in retrospective evaluations than others (e.g., Ariely & Carmon, 2000; Fredrickson & Kahneman, 1993). A review of these biases makes a memory-based approach very compelling in terms of developing our understanding of retrospective evaluations. Such an approach can easily and parsimoniously account for the larger impact of specific moments on retrospective evaluations simply by predicting their heightened accessibility in memory (Fredrickson, 2000; Tversky & Kahneman, 1973).

The studies reported below revisit this approach by providing a strong test of the role of memory in retrospective judgment. The task is simple:

participants are presented with a list of words and their task is to assess the pleasantness of the list; moreover all the experiments except the first, they also have to recall the items from each list that most easily come to mind. One important point to note about this setup is that as participants were told before hand about the evaluation task, on-line judgment formation was not inhibited. If the "on-line" judgment accounts are tenable, little or no association between memory and retrospective judgment should be observed as participants could entirely rely on the impression formed on-line. Secondly, the nature of the associations between memory and judgment was explored by manipulating the memorability of materials: first, one of the items was made distinctive; it was a negative stimulus amongst otherwise neutral materials; second, memory was manipulated through serial position effects (e.g., Tan & Ward, 2000). Based on these manipulations and a memory-based approach, we made predictions as to the changes in retrospective judgments that would appear in the various experimental conditions.

Finally, the studies reported below avoided the methodological pitfalls discussed previously. Firstly, we used a range of memory measures, including the ratio discussed above. Secondly, the design called upon included systematic controls over variables which could potentially confound any judgment and memory measurements. Hence, in the present series of studies, valenced and non-valenced words were selected as the components of to-be-evaluated lists, and length, familiarity and frequency of stimuli were all taken into account.

2.3. Experiment 1

The aim of the first study was to determine if the summary evaluations of word lists were susceptible to biases in the same way as other stimuli are; also, we wanted to examine if these biases would be the ones expected from the perspective of a memory-based approach. The rationale underlying the selection of words as the "units" of the to-be-assessed events (i.e. word lists) was threefold. First, extensive research has been conducted on memory for lists of words and mnemonic behaviour where they are concerned is well documented. Secondly, it was possible to exercise control over variables that affect judgment (i.e. valence), while at the same time controlling for other factors that are known to influence memory processes (i.e. stimulus length, frequency and familiarity).

Different predictions can be derived from the various approaches briefly reviewed in the introduction. A comparative assessment of our predictions and those of other views is provided in the general discussion. According to a memory-based approach, negative items presented in the first (primacy) and last (recency) positions of a list would have a memory advantage compared to negative items which are presented in the middle positions (e.g., Murdock, 1962). It follows that a memory-based approach predicts that lists with negative items presented in the first (primacy) or last (recency) positions would be rated as more unpleasant than lists where the negative items are presented in the middle positions. To our knowledge, no other theoretical account predicts both primacy *and* recency effects in retrospective judgment for the same stimuli – unless the nature of the task is manipulated (Hogarth & Einhorn, 1992; Zauberman, Diehl, & Ariely, 2006).

2.3.1. Method

2.3.1.1. Participants

A total of 21 participants (10 females) took part in this study; their ages ranged from 21 to 40 years (M = 27.9, SD = 4.9) and they were recruited via an e-mail advert.

2.3.1.2. Design and Materials

A pool of 192 words was selected from the Affective Norms of English Words database (ANEW; Bradley & Lang, 1999). Twenty-four negative items were selected along with 168 neutral ones. The selection was based on the valence (i.e. how positive or negative they are) and arousal scores (e.g., how much activation the normative sample reported on reading the word) of each item on database scales. Negative items were selected to be low in valence (less than 3, on a scale of 1-9) and high in arousal (over 5.9, on a scale of 1-9). Neutral items scored in the middle range for valence (4.5 to 6.9) and low on the arousal scale (less than 5). The selected negative items scored significantly lower than neutral items on valence scores (M = 2.3 and M = 5.7, respectively), $t(190)$ = 25.3, d = 6.1,[3] and had a significantly higher arousal rating (M = 6.5 and M = 4.1, respectively), $t(190)$ = 21.1, d = 4.8.

From the resulting word pool, 32 six-word lists were created, as follows. Eight lists included a negative item in the first position followed by 5 neutral words—hereafter identified as P1 lists. Four "P3" lists had a negative item in 3rd position while four "P4" lists had a negative item in 4th position. Eight "P6" lists

[3] Cohen's d is being used as the measure of effect size (Cohen, 1988)

comprised five neutral items and a negative word in last position. Finally, eight control lists contained only neutral words.

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 the Kucera-Francis frequency index (Kucera & Francis, 1967).

2.3.1.3. Procedure

After having agreed over e-mail to take part in the experiment, participants were provided with an application written in Authorware 7.0 (Adobe / Macromedia, 1987-2003). A series of introductory screens familiarised participants with the computer-controlled procedure and gathered demographic data.

Participants were told that the aim of the experiment was to collect normative data regarding the pleasantness of some word lists. They were instructed to attend to the lists and to provide an overall pleasantness rating for each one immediately after its presentation. The ratings were on a 0-100 scale (0 = *very unpleasant*, 100 = *very pleasant*), and participants were encouraged to make use of the whole range in their responses.

Each word of the 6-word lists was presented for one second with an inter-stimulus interval of 0.75 seconds. Immediately after the presentation of each list, participants were prompted to provide their rating, by typing in their numerical response. There was no time limit for the rating task. Participants also had the chance to become familiar with the procedure, through two practice trials.

In order to control for list presentation order, seven pre-defined versions of the experiment were prepared and randomly distributed to an equal number of participants. Each version included a pseudo-random order of the lists, since the randomisation process had one constraint: No more than three exemplars of each of the 5 list types (P1, P3, P4, P6 & Control) could occur consecutively.

Once the participants completed the experiment, they were instructed to return the results file to the experimenter via e-mail.

2.3.2. Results and Discussion

Figure 2.1 presents the mean pleasantness ratings as a function of list type. The mean pleasantness rating for the control lists (M = 68.3, SD = 10.0), with no negative item, was higher—that is more positive—than the ratings for the other list types. Moreover, there is some evidence of both primacy and recency in the ratings: P1 and P6 lists—which had a negative item in the first and last position respectively—received more negative ratings than the lists with a negative item in the middle positions.

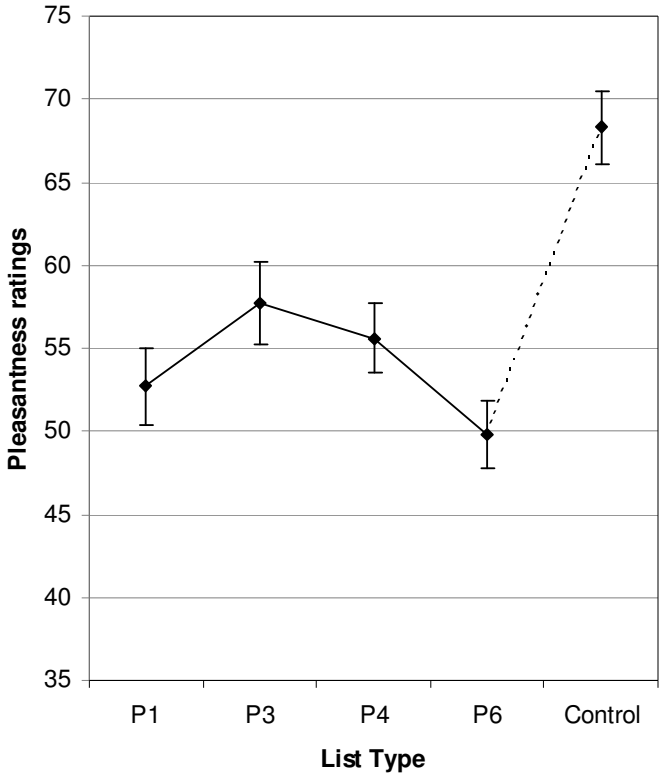


Figure 2.1 Mean pleasantness ratings as a function of list type. Error bars indicate *SEM* (Exp.1)

The inferential statistics confirmed these observations. Alpha was set to .05 for all analyses. A one-way ANOVA was run, with list type (P1, P3, P4, P6 and Control) as the within-subject factor. There was a significant main effect of list type, $F(4, 80) = 22.3$, $p < .001$, $\eta_p^2 = .53$.[4] Planned contrasts revealed that pleasantness ratings for P1, P3, P4 and P6 lists were significantly lower than for Control lists (all *p*s < .001). Conversely, judgments for P6 or recency lists were lower than those for P4 lists (pre-recency lists), $t(20) = 3.3$, $p < .01$, $d = .71$. Finally, ratings for P1 or primacy lists were lower than ratings for P3 lists, $t(20) = 2.1$, $p < .05$, $d = .46$.

[4] Partial Eta Squared is used as the measure of effect size

In sum, it can be concluded that the inclusion of a negative item affected perceived list pleasantness. More importantly, the negative item's effect was moderated by its presentation position, since a peak in first or last position exerted a larger impact on summary assessments than did negative items presented in comparison positions elsewhere in the lists.

When considering the judgment pattern, a straightforward interpretation can be derived from a memory-based approach. According to this view, the extent to which the negative item was available at the time of judgment would be the best predictor of its effect on retrospective evaluations. One of the most robust findings in the memory literature for lists of the type used here are primacy and recency effects (e.g., Tan & Ward, 2000), namely the enhanced memorability for items presented at the beginning and at the end of a series. Therefore, one interpretation of the reported results is that the primacy and recency effects in retrospective evaluations observed in Experiment 1 are linked to a memory advantage for the negative items which were presented either in first or last positions. Experiment 2 examined this hypothesis.

2.4. Experiment 2

The main objective of Experiment 2 was to introduce a memory measure into the retrospective evaluation task in order to more directly investigate the relationship between retrospective judgment and memory for each series. As in Experiment 1, primacy and recency effects in retrospective judgment are predicted in that lists with a negative item in either the first or last position are expected to be more negatively assessed than the other lists, the latter including

control lists and lists with negative items in the middle positions. At the same time it is expected that negative items presented either at the beginning or at the end of the sequence will be characterised by a memory advantage compared to distinctive items presented in the middle positions. Simply put, the prediction is that primacy and recency effects should be observed for the recall of the negative items.

The predictions of the memory-based approach are tested further through various means. First, correlations between memory and judgment were examined. In doing so, we compared different measures. We computed a global, ratio-type, memory measure, as this was called upon in previous studies and we have argued that it may have masked the relationship between memory and judgment (e.g., Hastie & Park, 1986; Lichtenstein & Srull, 1987). This memory ratio measure reflected the degree of isolation with which the negative item was recalled: the higher the ratio, the fewer neutral items recalled together with the negative word. We then correlated this measure of memory with a retrospective judgment measure.

Secondly, judgment was required first and memory for the content of the word-list obtained second (details of how this was done follow below). Hence, it was possible to contrast the mean pleasantness rating obtained when the negative item was recalled with the mean rating for the trials when it was not recalled. Presumably, if a distinctive item is available for later recall, it is more likely to have been 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, we would expect that on average, the pleasantness

rating will be lower in the cases where the negative item was available for the memory component of the task.

Thirdly, the effect of negative item availability was further analysed by examining ease of recall. As a measure of relative memory accessibility, we used output position in the memory task. Since participants were asked to perform free recall (hence no output constraints were implemented), we made the simplifying assumption that items recalled first are on average more readily accessible in memory. It was assumed that negative items recalled early on were more easily accessible than negative items that are recalled later on and would have had more impact on the retrospective assessment. Hence, our hypothesis was that the earlier a negative item was recalled, the stronger its impact on retrospective evaluations. For this reason, the effect of this differential accessibility of negative words was analysed in relation to the retrospective evaluations participants provide for the *lists as a whole*.

2.4.1. Method

2.4.1.1. Participants

Thirty-six undergraduate students (31 females) from City University London took part in the study. Age ranged from 18 to 56 years (M = 26.7, SD = 10.6). Participants were granted course credits for an introductory course in psychology for their participation.

2.4.1.2. Design and Materials

The materials employed were the same as in Experiment 1.

2.4.1.3. Procedure

All participants took part in an individual testing session that lasted approximately 45 minutes. The experiment was controlled by a computer program developed specifically for Experiment 2 with Authorware 7.0 (Adobe / Macromedia, 1987, 2003).

As in Experiment 1, participants were told that the aim of the experiment was to collect normative data about the pleasantness of word-lists. Word lists contained six words here also. The scale used for the participant ratings was the same as in Experiment 1. In this study, a series of asterisks appeared on the screen for 3 seconds to signal the end of the list presentation. Also, instead of typing in their responses, participants were required to use the mouse to click on a slide bar (with extremes of 0 and 100) on the position they felt was closest to their impression of the list. In order to limit the extent of anchoring effects (e.g., Chapman & Johnson, 2002) a sliding marker would appear on the bar (with its equivalent numerical value underneath) only after participants clicked for the first time on the slide bar. Participants then had the opportunity to adjust this initial rating by sliding the marker, and were to confirm their final one by clicking on a "Continue" button.

After rating a given list's overall pleasantness, participants were required to perform a free recall task, i.e. they were to type all the words they could remember from the most recently presented list. The instructions emphasised that spelling errors would not affect scoring and that both the assessment and recall

tasks were equally important; participants were asked not to overlook the rating task in order to proceed more quickly to the recall task. After participants had completed the recall task, the following list was presented, and so on.

Three practice trials were provided. List presentation order was randomised independently for each participant and no time limits were set for the rating and recall tasks. A post-experimental questionnaire was used at the end of the session in order to gather information about how participants completed both the judgment and recall tasks.

2.4.2. Results and Discussion

Because of 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 in order to proceed more quickly to the memory task. Participants whose judgment scores were characterised by a standard deviation of 5 or less (5% of the scale) were eliminated from the analyses. One participant was excluded according to this criterion. Moreover, in the post-experimental questionnaire, the same participant indicated that s/he performed the pleasantness ratings according to list memorability rather than perceived pleasantness. Alpha was set to .05 for all analyses.

Judgment. Figure 2.2 presents the mean pleasantness ratings as a function of list type. A perusal of Figure 2 shows a pattern of results that is very similar to the one obtained for Experiment 1. First, pleasantness ratings for the Control lists were the highest (M = 60.8, SD = 12.4). Moreover, lists with a negative item in

either first (M = 44.3, SD = 13.3) or last position (M = 44.2, SD = 12.5) were rated as more unpleasant than lists with a negative item in either 3rd or 4th positions (M = 49.3, SD = 13.3; M = 49.4, SD = 12.8, respectively).

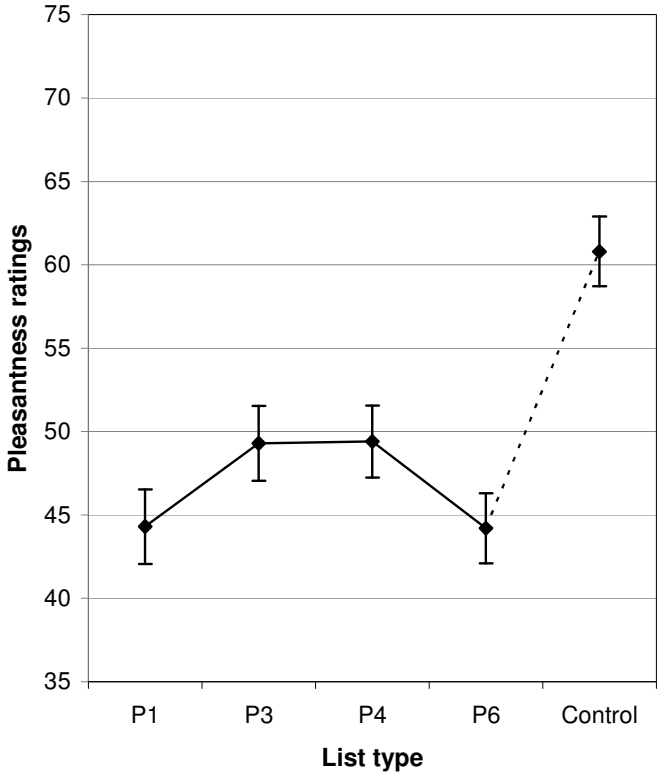


Figure 2.2 Mean pleasantness ratings as a function of list type. Error bars indicate *SEM* (Exp.2)

A one-way ANOVA was run, with list type (P1, P3, P4, P6 and Control) as the within-subject factor. Overall, there was a significant main effect of list type, $F(4, 136) = 35.9$, $p < .001$, $\eta_p^2 = .52$. Planned contrasts indicated that pleasantness ratings for P1, P3, P4, and P6 lists were significantly lower than for Control lists (all ps < .001).

Planned contrasts also revealed that the pleasantness ratings for P1 (primacy) lists were lower than for P3 lists, $t(34) = 3.9$, $p < .001$, $d = .65$. Moreover, a recency effect was observed too as the ratings for P6 lists were lower than for P4 lists, $t(34) = 4.3$, $p < .001$, $d = .73$.

Memory. Figure 2.3 represents the mean recall proportion for the items as a function of word position and valence. It seems that overall there was a memory advantage for the negative items as compared to the neutral ones presented in the same position; participants exhibited higher recall rates for the Peaks than for the neutral items, across presentation positions. More importantly, memory primacy and recency effects for the negative items can be observed: Peaks presented in either first ($M = .78$, $SD = .17$) or last position ($M = .71$, $SD = .23$) were better recalled than Peaks presented in either 3rd or 4th positions respectively ($M = .56$, $SD = .28$; $M = .56$, $SD = .30$).

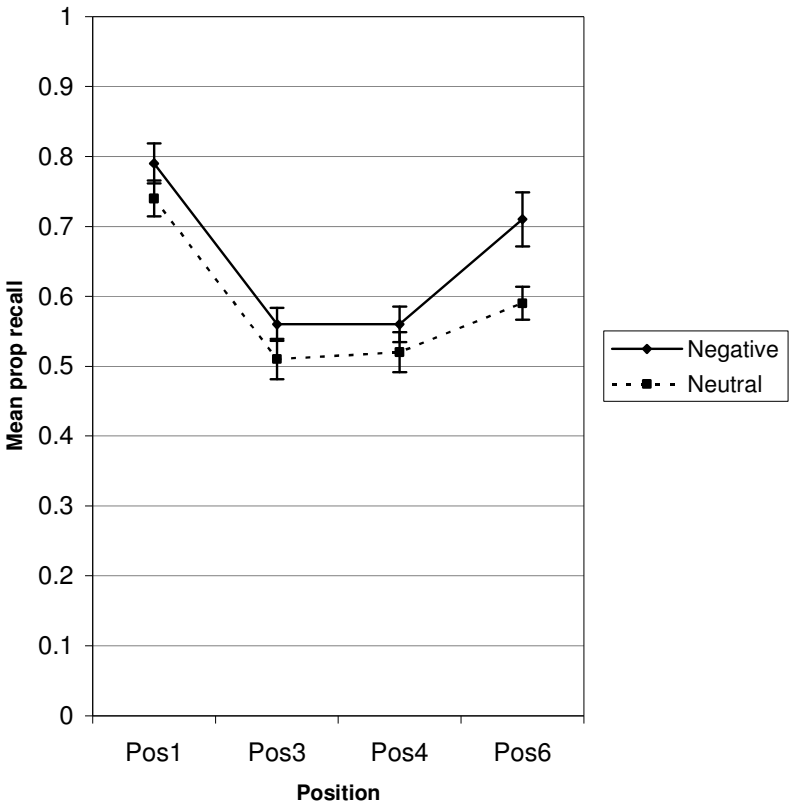


Figure 2.3 Mean recall as a function of word position and valence. Error bars indicate *SEM* (Exp.2)

The data were analysed using a 2 (valence: negative vs. neutral) × 4 (position: 1st, 3rd, 4th, and 6th) repeated measures ANOVA. Main effects of position ($F(3, 102) = 22.2$, $p < .001$, $\eta_p^2 = .40$) and valence ($F(1, 34) = 8.8$, $p < .01$, $\eta_p^2 = .21$) were noted. The word valence by position interaction was not significant ($F < 1$), indicating that the memory advantage for the "Peaks" over the corresponding neutral items was relatively constant across positions.

To test for primacy and recency effects in the recall of the negative items, planned contrasts analyses were computed. The results showed that P1 recall was higher than P3 ($t(34) = 5.4$, $p < .001$, $d = .92$) and also that P6 recall was higher

than P4 ($t(34) = 2.6$, $p < .05$, $d = .44$). Hence, primacy and recency effects were observed for the recall of the negative items.

Memory-Judgment relationships. In order to reduce the influence of potential anchoring effects (e.g., Chapman & Johnson, 2002) and of inter-individual differences in the use of the 0-100 scale, judgment scores were transformed as follows: For each participant, the average pleasantness rating for the Control lists was subtracted from the pleasantness ratings for each P1, P3, P4 and P6 list – that is the lists that contained a negative item. The new corrected judgment scores (J') therefore represented how much more unpleasant each P1, P3, P4, and P6 list was in comparison to the average Control list for each participant. J' scores were then averaged for each participant, according to the negative item presentation position and whether the negative item presented in the list was recalled or not.[5]

Then a "global" memory score for each participant was computed (e.g., Hastie & Park, 1986). This score was the average ratio between the negative information recalled and the total amount of words recalled: the higher the value, the more *negative* the memory for the list. Thus, the lowest possible score of 0 corresponds to neutral items being recalled and the maximum score of 1 refers to lists where only the negative item was recalled. We then correlated this measure with the overall average corrected judgment (J') for each participant: It was expected that the more negative the memory for the list, the lower the pleasantness ratings. However, the correlation yielded non significant results, Spearman's ρ (33) = -.118, $p > .49$, and revealed how memory and judgment

[5] This analysis yielded a total of 9.6% missing values. 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, we will be reporting the data obtained via mean by subject substitution.

measures were not associated – at least when using this ratio-style memory measure.

Second, we computed more comprehensive memory measures: we compared the corrected average pleasantness rating for lists where the negative item was recalled versus lists where the negative item was not recalled. Overall, when the negative item was recalled in the memory task, pleasantness ratings were lower (M = -15.7, SD = 11.3) than when the negative item was not recalled (M = -9.9, SD = 9.3). When these results were broken down by list-type, the same pattern appeared for P1 and P4 and P6 lists. For P3 lists, the pleasantness ratings seem low regardless of whether the negative item was recalled or not.

A 2 (Memory: Negative item recalled Vs. not recalled) by 4 (List type: P1, P3, P4 and P6) within-subjects ANOVA confirmed these observations. The main effect of Memory was significant ($F(1, 34) = 16.7$, $p < .001$, $\eta_p^2 = .33$), confirming that overall ratings were more unpleasant for those lists where the negative word was recalled. The List by Memory interaction was significant, too ($F(3, 102) = 5.5$, $p < .01$, $\eta_p^2 = .14$). Follow-up analyses revealed that the main effect of Memory was significant for P1 and P4 lists ($t(34) = 4.5$, $p < .001$, d = .75; $t(34) = 3.2$ $p < .01$, $d = .72$, respectively). However, for the P3 and P6 lists the pleasantness ratings were invariably low regardless of the negative item being recalled or not (both ps > .05). Table 2.1 below summarises these findings.

Table 2.1 Mean corrected pleasantness ratings (J') as a function of list type and "Peak" being recalled or not (Exp.2)

		List type			
		P1	P3	P4	P6
Was the Peak recalled?					
No	*M*	-7.5	-11.9	-5.5	-14.8
	SD	(12.7)	(13.2)	(11.8)	(14.9)
Yes	*M*	-19.1	-11.8	-14.7	-17.2
	SD	(13.3)	(13.4)	(13.4)	(11.0)

Finally, retrospective evaluations were analysed depending on the negative item recall position (see Table 2.2). In other words, the pleasantness ratings (J') were examined according to the position in which the negative item was recalled by the participants (regardless of its presentation position). The underlying rationale was that items that are more accessible in memory are likely to be recalled earlier—if the negative item is more accessible and recalled early we would expect its impact on retrospective evaluations to be higher than when it is recalled later in the protocol or not at all.

Table 2.2 Mean corrected pleasantness ratings (J') as a function of "Peak" recall output position (Exp.2)

		Peak recall output position		
		Not recalled	Recalled in pos 1,2	Recalled in pos 3-6
Pleasantness ratings (J')	*M*	-10.3	-18.9	-13.8
	SD	(10.7)	(11.7)	(10.8)

Overall, the pleasantness ratings varied depending on the negative item output position: They were lowest when the participants recalled the negative item as either the first or second response (M = -18.9, SD = 11.7) and they were the highest when participants did not recall the Peak at all (M = -10.3, SD = 10.7). The lists were rated somewhere at an intermediate level when the negative item was recalled as either 3rd, 4th, 5th or 6th response (M = -13.8, SD = 10.8).

A significant main effect of recall position on pleasantness ratings ($F(2, 68) = 16.0$, $p < .001$, $\eta_p^2 = .32$) confirmed these observations. A planned contrast analysis revealed 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, $t(34) = 4.3$, $p < .001$, $d = .72$, which in turn were lower than for those lists where the Peak was not recalled at all, $t(34) = 2.3$, $p < .05$, $d = .38$.

An overview of the results of this second experiment indicates that they confirm the pattern observed in Experiment 1. Once again, a negative item presented either at the beginning or at the end of the series exerted a larger impact on pleasantness evaluations.

Furthermore, the recall results were as expected from the perspective of a memory-based approach: Negative items were better remembered than corresponding neutral items, across presentation positions. More importantly, the memory pattern for the negative items displayed both primacy and recency effects, providing support to the possible contribution of memory processes to judgment biases in retrospective evaluations.

The results of the correlational analyses produced different outcomes, depending on the memory measure that was being used. When a global, ratio-style, memory measure was utilised (e.g., Hastie & Park; Lichtenstein & Srull,

1987) – where each recalled item holds the same weight – no significant associations were observed between memory and judgment. However, when the role of memory was tested through more comprehensive measures – or measures which took item accessibility into consideration also (e.g., Schwarz, 1995) – the results provided support for a memory-based approach. Pleasantness ratings for those lists where the negative item was recalled were significantly lower than for those lists where the negative item was not recalled. This result supports the idea that when a negative item was easily available in memory at the time of judgment it exerted a higher impact on judgment. The assumption was that when a negative item was not recalled in the memory task, it was also less likely to be available at the time of retrospective evaluations; on average, this would lead to a less negative assessment of the list.

These results suggest that participants relied at least to some extent on episodic information stored in memory when providing retrospective evaluations. If they had exclusively relied on on-line judgment formation, there would be no reason to expect an association between memory and judgment measures, although as is always the case with a correlational approach a third factor could perhaps be causing changes in both memory and retrospective evaluations. However, accessibility in memory of a negative item seemed to moderate retrospective judgment, since lower ratings were associated with the negative item being recalled early in the response sequence. The on-line view does not lead to the expectation that the accessibility of the distinctive-negative item would have an impact on retrospective evaluation.

Some of the more detailed follow-up analyses showed how the above mentioned association between memory and judgment was not observed when

the negative item was presented in either 3rd or last position. This absence of relationship is of course potentially attributable to the reduction in power associated with the more detailed breakdown of data. In order to assess whether this finding was a reliable one or not, a further experiment was conducted. Also, Experiment 3 involved a change in the memory task. In Experiment 2, participants were required to remember as many words as they could from each list, in any order. Previous research has shown that people rarely retrieve all the information that is available in order to perform a judgment task. It seems instead that people "…truncate their search process as soon as enough information has come to mind to form a judgment with sufficient subjective certainty" (Schwarz & Vaughn, 2002; p.105). Generally speaking, it seems reasonable to argue that typical retrospective evaluations do not place high demands on memory precision. It follows that in a scenario more closely related to typical retrospective evaluations situations participants would not be asked to recall all the information they can about a specific experience after having assessed it in hindsight.

Moreover, in Experiment 2, there were some concerns about the effect of the demands of the memory task—as list length exceeded memory capacity, participants had less than perfect performance. It was not clear what effect the difficulty of the free recall task was having on the retrospective evaluations component of the Experiment. One criticism could perhaps be that the difficulty of the memory task biased the retrospective evaluations in such a way as to artificially inflate the relationship between the summary assessments and the memory measures. Experiment 3 revisited the memory-assessment relationship while addressing the issues above.

2.5. Experiment 3

In this study, in order to reduce the cognitive demands of the memory task, participants were asked to recall only the first two items that came to mind from the just-seen list, instead of performing a free recall task requiring the recall of all the presented words. Hence, this experiment investigated a situation where the retrospective evaluations were coupled with a much less demanding memory task: Recalling two items from a six-item list is well within the memory capacity of normal adult participants and can be done without any deliberate effort to memorise the presented items.

2.5.1. Method

2.5.1.1. Participants

A total of 38 participants (21 females) took part in the study. Ages ranged from 19 to 55 years (M = 40.9, SD = 11.4). Participants were recruited through local advertising and were rewarded with £7 per hour in exchange of their participation.

2.5.1.2. Design and Materials

The materials selection process changed slightly as compared to Experiments 1 and 2. The same pool of 192 words was employed, but new lists were generated and a few further controls were implemented. Through within-list

matching between negative and neutral items, 32 six-word lists were constructed as before. The lists were then assigned to 4 different blocks of 8 lists each, with blocks A, B and C including "negative" lists (which included 5 neutral and 1 negative items) and block D consisting of the "Control" lists (6 neutral items). The four blocks were closely matched on familiarity ratings (Coltheart, 1981), with block averages ranging from 525.7 to 536.1 (scale from 100 to 700), number of phonemes (from 4.3 to 5) and the Kucera-Francis frequency index (Kucera & Francis, 1967), with blocks averages ranging from 37.2 to 41.2 (scale from 1 to 10,000).

Moreover, negative items were rotated, so that each of the 24 negative words selected from the database (ANEW; Bradley & Lang, 1999) appeared in each possible position (1st, 3rd, 4th and 6th) an equal number of times across participants.

2.5.1.3. Procedure

Each participant took part in an individual testing session that lasted approximately 40 minutes. The whole experiment was run on Authorware 7.0 (Macromedia, 1987-2003). The overall procedure was identical to Experiment 2, the only difference being the nature of the memory task. After providing an overall pleasantness rating, participants were asked to type the first two words that came to mind from the just-seen list, and then to click the "Continue" button. On the following screen they had the chance to type in any additional words they might remember, but it was stressed that it was equally fine to proceed directly to the next trial.

2.5.2. Results and Discussion

Two participants were excluded from the analyses: One participant rated list memorability instead of pleasantness, while the other participant's standard deviation for the pleasantness ratings across trials was equal to 3.4, hence lower than the a priori determined cut-off point of 5 (see Experiment 2). Alpha was set to .05 for all analyses.

Judgment. Figure 2.4 presents the mean pleasantness ratings as a function of list type. In accord with the two previous experiments, Control lists were rated as the most pleasant lists (M = 65.0, SD = 14.4). Once again, primacy (P1, M = 48.9, SD = 16.8) and recency lists (P6, M = 50.6, SD = 17.1) were rated as more unpleasant than middle lists (P3, M = 53.1, SD = 16.1; P4, M = 54.6, SD = 17.3).

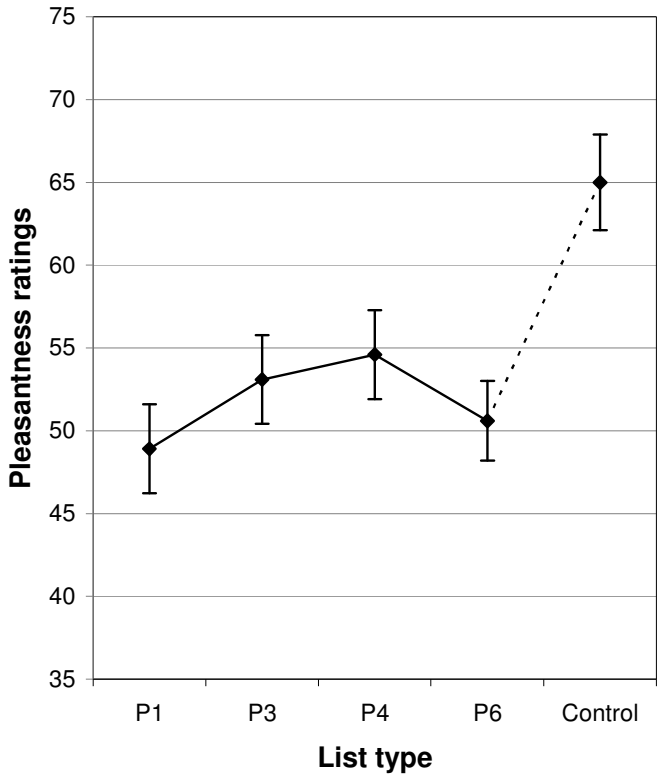


Figure 2.4 Mean pleasantness ratings as a function of list type. Error bars indicate *SEM* (Exp.3)

Analyses were conducted with a one-way ANOVA, with list type (P1, P3, P4, P6 and Control) as the within-subject factor. In line with Experiments 1 and 2, there was a significant main effect of list type, $F(4, 140) = 22.8$, $p < .001$, $\eta_p^2 = .41$. The pleasantness ratings for the Control lists were significantly higher than for P1, P3, P4 and P6 lists (all *p*s < .001). Once again, P1/ primacy lists were rated as more unpleasant than P3 lists, $t(35) = 2.1$, $p < .05$, $d = .36$. Moreover, recency effects were again observed, as pleasantness ratings were lower for P6 lists as compared to P4 lists, $t(35) = 2.2$, $p < .05$, $d = .35$.

Memory. First, let us briefly comment on the effect of the change in memory task. For each participant we computed the proportion of trials (out of 32) where

they recalled two or fewer items. A t-test was carried out, with the proportion as the dependent variable and Recall task (Exp 2, free recall and Exp 3, two-item recall) as the between-subjects factor. The t-test was significant, $t(69) = 2.2$, $p < .05$, $d = .52$, and it confirmed that participants were more likely to report two or fewer items in the current experiment ($M = .28$) – relative to Experiment 2 ($M = .19$).

Secondly, memory performance was analysed depending on item valence and presentation position. In line with Experiment 2, memory was superior for negative items than for neutral ones, across positions. Moreover, participants more often recalled the negative items presented in either the first (primacy, $M = .77$, $SD = .20$) or the last positions (recency, $M = .63$, $SD = .23$), as compared to the middle ones (P3, $M = .60$, $SD = .31$; P4, $M = .51$, $SD = .31$). Figure 2.5 represents the mean recall proportion as a function of word position and valence.

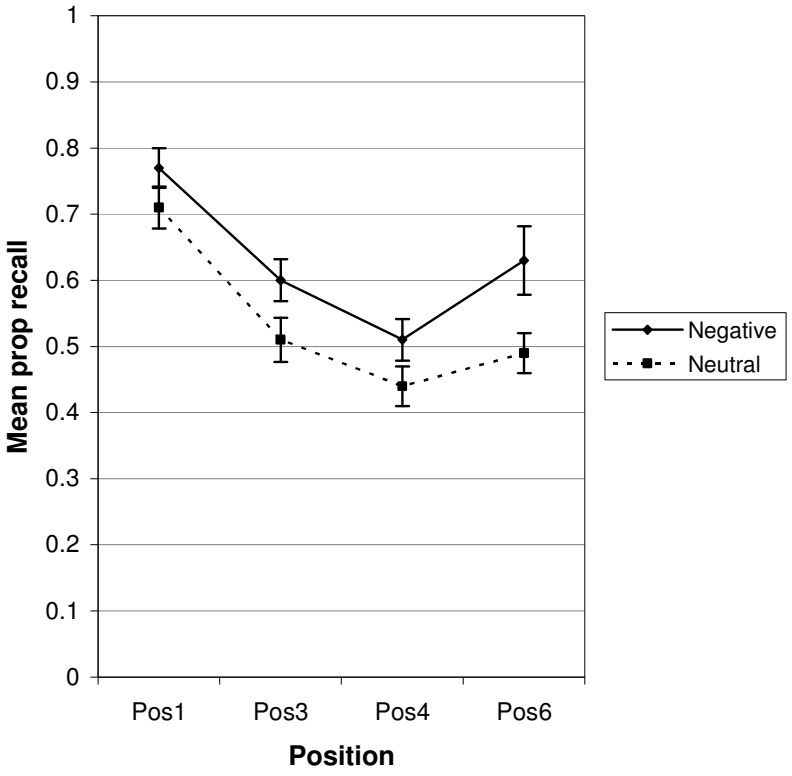


Figure 2.5 Mean recall as a function of word position and valence. Error bars indicate *SEM* (Exp.3)

These data were analysed using a 2 (valence: negative vs. neutral) X 4 (position: 1st, 3rd, 4th and 6th) repeated measures ANOVA. Corroborating the findings in the previous two experiments, the main effects of position ($F(3, 105) = 17.3$, $p < .001$, $\eta_p^2 = .33$) and valence ($F(1, 35) = 11.0$, $p < .01$, $\eta_p^2 = .24$) were significant. As before, the word valence by position interaction was not significant, $F(3, 105) = 1.0$, ns. Planned contrasts analyses confirmed that primacy and recency effects were observed for the recall of the negative items. When presented in the first position they were better remembered than when presented in 3rd position, $t(35) = 3.6$, $p < .01$, $d = .59$. Moreover, recall of

negative items in last position was higher than recall of negative words in 4th position, $t(35) = 2.1$, $p < .05$, $d = .35$.

Memory-Judgment relationships. As in Experiment 2, judgment scores were transformed; once again, the corrected judgment score (J') for each list represented how much more unpleasant (or pleasant) that specific list was compared to the average rating that each participant assigned to the Control lists. As in Experiment 2, analyses were performed on the corrected scores, comparing mean judgment scores depending on whether the negative item was recalled or not.[6]

First, as in Experiment 2, a global, ratio-style memory measure was computed for each participant: the higher the ratio, the more negative the memory for the list. This memory measure was then correlated with the average corrected pleasantness rating for each participant. Contrary to Experiment 2, this correlation was significant, Spearman's ρ (34) = -.351, $p < .05$. This preliminary analysis suggests that, as memory task demands are lowered, the relationship between memory and judgment is strong enough to show up with a global ratio-like memory measure.

Then, as in Experiment 2, further analyses were run to investigate the nature of the relationships between memory and judgment. Overall, pleasantness ratings again were lower for those lists where the negative item was recalled (M = -15.7, SD = 10.7) as compared to lists where it was not (M = -8.5, SD = 9.2).

[6] As for Experiment 2, this analysis yielded missing values (10.4% of the total cells). Different strategies were again called upon for dealing with these missing values - including mean substitution by subject, grand mean, and Expectation-Maximization algorithm (Schafer & Olsen, 1998). As all the analyses returned the same results, we will be reporting the data obtained via mean by subject substitution.

Contrary to Experiment 2, this relationship between memory and judgment was evident across list types.

A 2 × 4 repeated-measures ANOVA was run with Memory (negative item recalled Vs. not recalled) and List type (P1, P3, P4 and P6) as factors. The analyses yielded a significant main effect of Memory ($F(1, 35) = 15.3$, $p < .001$, $\eta_p^2 = .30$), confirming that overall pleasantness ratings were significantly lower for those lists where the negative word was recalled. In contrast to Experiment 2, the relationship between memory and judgment was observed regardless of presentation position, as the List by Memory interaction was not significant ($F(3, 105) = 1.4$, $p > .24$). Table 2.3 below summarises these findings.

Table 2.3 Mean corrected pleasantness ratings (J') as a function of list type and "Peak" being recalled or not (Exp.3)

		List type			
		P1	P3	P4	P6
Was the Peak recalled?					
No	*M*	-7.7	-8.5	-6.8	-11.1
	SD	(12.1)	(14.8)	(12.3)	(12.5)
Yes	*M*	-18.3	-13.7	-15.2	-15.6
	SD	(13.0)	(15.9)	(14.8)	(11.4)

As a final analysis, retrospective evaluations for all list types were analysed depending on the negative item recall or output position, regardless of its presentation position (see Table 2.4). Ratings were averaged across list types

and compared depending on the position in which the participant recalled the negative item.[7]

An examination of the means shows that, overall, the most unpleasant ratings were provided for those lists where the negative item was recalled amongst the first two responses (M = -19.6, SD = 12.5), followed by the ratings for those lists where the negative item was recalled as either the 3rd, 4th, 5th or 6th response (M = -11.6, SD = 9.4). Once again, the least unpleasant ratings were provided for those lists for which the negative word was not recalled at all (M = -7.6, SD = 8.5).

In accord with Experiment 2, the main effect of negative item recall position ($F(2, 70) = 21.8$, $p < .001$, $\eta_p^2 = .38$) was significant. A planned contrast analysis revealed once again that average judgment scores were lower when the negative word was recalled either first or second compared to when it was recalled as either 3rd, 4th, 5th or 6th response, $t(35) = 5.0$, $p < .001$, $d = .83$. Finally, when the negative word was not recalled ratings were more pleasant than when it was recalled among the last 4 responses, $t(35) = 2.4$, $p < .05$, $d = .40$. Hence, here also, the findings suggested that if the negative item was more easily accessible its impact on retrospective evaluation was larger than when it was less accessible or not retrievable.

[7] This analysis yielded a total of 2.8% missing values, which were once again imputed via mean by subject substitution.

Table 2.4 Mean corrected pleasantness ratings (J') as a function of "Peak" recall output position (Exp.3)

		Peak recall output position		
		Not recalled	Recalled in pos 1,2	Recalled in pos 3-6
Pleasantness ratings (J')	*M*	-7.6	-19.6	-11.6
	SD	(8.5)	(12.5)	(9.4)

In summary, these results reproduced the judgment score patterns observed for the previous two experiments: Both primacy and recency were observed in retrospective evaluations. If a negative item was presented either in the first or last positions, its effect on the pleasantness ratings was higher than if it was presented in the middle positions. This pattern was once again mirrored in the memory participants displayed for the negative items, which were better remembered if presented in first or last positions. Also, the effect of a negative stimulus on retrospective evaluations could be predicted by its recall position: the more accessible in memory an item was, the larger its impact on summary assessments.

Experiment 3 also suggested that lowering the memory task demands affected the relationship between memory and judgment relative to Experiment 2. More specifically, memory-judgment correlations were observed irrespective of the presentation position of the distinctive information, thereby providing stronger support for memory-based judgment in retrospective evaluations.

2.6. General Discussion

The general objective of the present paper was to investigate the contribution of memory processes to judgment biases in retrospective evaluations. A memory-based approach was put forward and its predictions systematically tested. We would argue that this constitutes a strong test of the influence of memory on retrospective judgment, as the latter was successfully predicted on the basis of the former. Relative to previous research that focused on the relationships between memory and judgment (e.g., Anderson & Hubert, 1963; Dreben et al., 1979), the present study implemented more stringent controls over variables that are known to influence memory performance; also, more comprehensive memory measures were utilised. By doing so, a more thorough investigation into the role of memory in retrospective judgment could be carried out and the results could serve to make sense of contradictory results reported in the literature in the area.

The pattern of results for summary assessments of lists containing a distinctive negative item was largely constant across experiments. More specifically, the position in which a distinctive negative stimulus was presented exerted differential effects on judgment; negative items presented either in the first or last position of a list affected overall pleasantness ratings to a greater extent than negative items presented in the middle positions. Simply put, primacy and recency biases were observed for retrospective evaluations of sequences containing a distinctive event.

Such findings are somewhat at odds with the predictions that can be derived from the Peak-End rule (Fredrickson & Kahneman, 1993). Based on this

heuristic, it would be reasonable to expect that the effect of a negative item on pleasantness ratings should be the greatest when the "Peak" was presented last: In these situations, the "Peak" and the "End" would overlap; hence, no averaging would be required in order to best predict the summary evaluations. When the negative item was not presented in last position, the averaging predicted by the Peak-End rule would involve the last item (a neutral word) and the Peak item (the negative word), and no differences in the evaluations should stem from the negative item being presented in first or middle positions[8].

At the same time, simple summative approaches (e.g., the "value-account"; Betsch et al., 2001) would predict neither primacy nor recency effects in retrospective evaluations: The decisive factor for summary evaluations is the quality of each piece of information, and not its position within the sequence. Hence, these approaches cannot account for the present array of results.

Finally, the fact that clear associations between memory and judgment were observed is contrary to the predictions of approaches that postulate functional independence between memory and judgment (e.g., Anderson, 1981; Betsch et al., 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. Hence, according to on-line approaches to summary evaluations, no correlations should have been observed between the overall lists pleasantness ratings and the memory performance participants exhibited for the items included in the lists. Moreover, according to Hastie and Park (1986), memory-judgment correlations should arise only when participants are not aware

[8] It could be argued that the nature of the stimuli considered in the present experiments vary from the majority of studies finding Peak-End effects (e.g., Kahneman et al., 1993). However, recent evidence has suggested that the Peak-End rule may also apply to discrete and semantic stimuli (e.g., Do et al., 2008; Langer et al., 2005) and not only to continuous and experiential stimuli.

of the subsequent judgment task – condition that does not apply to the present paradigm, as participants were told about the rating task beforehand.

Considering the memory performance for the negative items, it can be argued that a memory-based approach was the most effective in explaining the judgment biases observed. As expected, primacy and recency effects were observed for memory for the negative stimuli: when the negative item was presented in the first or last position, its recall rate was higher than when it was presented in the middle positions. This memory advantage was mirrored in the retrospective evaluations of the presented lists: Lists with negative items at the start or end were associated with a more negative retrospective evaluation.

In order to explore this hypothesis further, correlational analyses were performed and provided findings which are in accordance with the availability heuristic (Tversky & Kahneman, 1973). 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. This relationship between memory and judgment was even clearer when the memory task demands were lowered (Experiment 3), hence providing a more naturalistic setting for the global summary evaluations task.

Taken together, the results suggest that participants consulted the episodic record they retained from each list in order to assess it in hindsight. All of the findings reported in the three experiments are consistent with a memory-based approach where participants, prior to performing the pleasantness evaluation of each list, "looked back" at the remembered representation of each list, and based their retrospective evaluations on the information available.

A further finding supported the memory-based approach. When the negative information was more easily accessible in memory (as it was recalled amongst early responses), the summary evaluations were more affected by it. This pattern suggests that participants, when providing retrospective evaluation of list pleasantness, overweighed information from the just-presented list that was more easily accessible. This would in turn support the notion that availability – considered as ease of recall (Schwarz & Vaughn, 2002) – is a mediator of the impact that the encoded information will have on summary assessments.

However, it is perhaps important to note that the results suggest that the role of on-line impression formation (e.g., Hastie & Park, 1986; Hogarth & Einhorn, 1992) cannot be completely ruled out. It is noteworthy that overall pleasantness ratings were still affected even if the negative item was not produced at the recall phase. Pleasantness ratings for those lists where the negative item was not recalled (M = -9.9 and M = -8.5 for Experiment 2 and 3, respectively) were still lower than for those lists where no negative item was presented at all (the "Control" lists).

While it is possible that the negative item, accessed at the time of judgment production, was not available for recall later on because of situational factors (e.g., output interference; Nairne, 1990), it could also be hypothesised that the negative item affected the on-line judgment formation, and that its trace was not accessible at the time of judgment as such. One interpretation of these findings would be that the summary evaluations were the result of on-line impression formation – that served as a form of anchor – which was adjusted at the point of the actual judgment response depending on the information that was

most easily available and accessed. Such an explanation could go some way toward clarifying why correlations between memory and judgment have been difficult to obtain in previous studies in the social cognition literature (e.g., Hastie & Park, 1986). Most studies adopted a ratio as the memory measure (e.g., between positive and negative information), which assigns the same weight to any piece of information that is retrieved. A ratio measure does not include any information about the ease of recall of each piece of evidence, and therefore may not be an accurate measure of the representation people are accessing in order to evaluate an episode (or at least not powerful enough to detect the underlying memory contribution to judgment in hindsight).

In conclusion, by devising a paradigm that could systematically test for primacy, recency and general position effects on retrospective evaluations it was shown that distinctive moments of an experience impact summary assessments in a way that is predicted by a memory-based approach to biases in retrospective evaluations. The reported results provide compelling evidence in favour of a viewpoint where remembered episodic information has a biasing effect on judgment in hindsight.

Chapter 3:

The Effects of Manipulating Accessibility

3.1. Introduction

How do people provide overall assessments about affective episodes? Do they rely on information stored in memory about the event itself? In Chapter 2, clear associations between memory and judgment were obtained: primacy and recency were observed for both recall and retrospective judgment measures. Three further findings provided evidence in favour of the claim that memory and judgment processes might be functionally associated. First, evaluations of the lists were more unpleasant when the negative item was recalled. Second, when the negative information was most likely to be easily accessible – as it was recalled as an early response – the retrospective evaluations (hereafter RE) were the most unpleasant. Third, memory task demands seemed to moderate the relationships between memory and judgment. The association between the two processes were clearest when the memory task demands were the lowest. In line with previous findings that showed how rarely people retrieve all the information about an event in order to evaluate it (e.g., Higgins, 1996), the results from a simpler memory task resulted in the strongest correlations between memory and judgment.

Thus, Chapter 2 provided some evidence in favour of approaches that suggest a direct relationship between memory and judgment processes (e.g., Tversky & Kahneman, 1973; Dougherty, Gettys, & Ogden, 1999). As participants were told beforehand about the evaluation task, they could have relied on the impression about the lists formed on-line (e.g., Anderson & Hubert, 1963; Betsch, Plessner, Schwieren, & Gütig, 2001) – and hence, there would be no reason to predict associations between memory and judgment measures (Hastie & Park, 1986). Some authors have indeed concluded that the only

situations in which summary assessments are based on information available in memory are those where people cannot form an on-line impression of the to-be-evaluated event (e.g., Custer & Aarts, 2003; Hastie & Park, 1986; McConnell, Sherman, & Hamilton, 1994).

The aim of the research included in this chapter was to further explore the nature of the memory-judgment associations previously observed. After all, simple associations between memory and judgment measures do not preclude the possibility that these are in attributable to a third factor. For instance, Anderson (1989; see also Moser, 1992) puts forward that, even if the memory and judgment systems are functionally independent, "…they will generally be correlated because both depend on the same given stimuli." (p. 209). Further, the author states that memory-judgment correlations may be predicted by on-line judgment models, too: this is "…simply because some stimulus variables affect both forms of memory in the same way." (p. 209).

In order to provide clearer evidence that accessibility moderates retrospective judgment, one strategy would be to manipulate memorability of segments within the episodes and predict RE accordingly. Accessibility refers to the ease with which information is accessed in memory at the time of judgment (e.g., Schwarz, 1995; Schwarz, Bless, Wanke, & Winkielman, 2003; Tversky & Kahneman, 1973). To our knowledge, few researchers have attempted to manipulate memory in order to predict judgment (e.g., Gabrielcick & Fazio, 1984; Hanita, Gavanski, & Fazio, 1997; Lewandowsky & Smith, 1983). Also, differences between our paradigm and previous methods suggest that such an investigation would be of interest. For instance, Lewandowsky and Smith (1983) successfully predicted summary assessments depending on the memory manipulation they

implemented; however, the evaluation task was frequency estimation rather than retrospective evaluation (see also Gabrielcick & Fazio, 1984 – who used priming instead of repetition as the way to manipulate the memorability of specific items). Frequency estimations are likely to be ruled by processes and dynamics different from those usually called upon in qualitative assessments (see Hogarth & Einhorn, 1992).

Hence, in the present chapter we will manipulate memorability through two well established memory phenomena – the modality effect (Experiment 4) and the Von Restorff effect (Experiment 5). The predictions as to the effect of these manipulations on RE are derived from a memory-based approach and contrasted with the prediction derived from other views.

As a final aim, the research presented in this chapter will also address another issue that could change the interpretation of the findings reported in Chapter 2. In Experiments 2 and 3, participants were told in advance about the subsequent memory task. This awareness could have artificially inflated the correlations between memory and judgment measures. For example, participants could have adopted a strategy where they accessed the list representation in memory at the time of judgment in preparation for the upcoming memory task.

In order to rule out the above mentioned possibility, participants in Experiment 6 were asked to perform a *surprise* memory task: after having provided only RE for a number of trials, they were asked to perform a memory task after the last presented list. It is argued that, if memory-judgment associations are observed in this case as well, they are unlikely to be due to participants anticipating the memory task. This would allow us to rule out the

idea that the nature of the task called upon in the previous chapter artificially inflated the correlations between memory and judgment.

3.2. Experiment 4

In this experiment, the objective was to further test the hypothesis that retrospective evaluations are biased by retrieving information about the assessed episode. This was done by increasing the accessibility (e.g., Schwarz, 1995) of negative items within the lists in some trials. If people rely, at least in part, on the information retrieved about an event in order to evaluate it in hindsight, then it is reasonable to predict that instances where the negative information is more easily accessible will be rated as more unpleasant as compared to situations where the memorability of the negative information was not enhanced.

The memory phenomenon that was selected in order to manipulate item accessibility was the modality effect (e.g., Corballis, 1966; Murray, 1966). The modality effect refers to the better recall of items in the auditory modality relative to the same items presented in the visual modality. In effect, this auditory advantage is due to the better recall of the recency items – i.e. the items presented in the last positions: Lists presented aurally generate a significantly larger recency effect Earlier items are recalled at a similar level, regardless of modality of presentation. These findings are rather reliable; they have been replicated with a variety of stimuli and with different memory tasks (e.g., Crowder, 1986; Glenberg, 1984; Neath, 1997)

As in Experiments 1 to 3, Experiment 4 involved presenting lists of words that participants had to assess for pleasantness. The rationale supporting this choice of material was again twofold: first, it was possible to match negative

(distinctive) and neutral (background) items along features that could be affecting memory at the retrieval stage, hence potentially distorting the underlying relationships between memory and judgment measures. Secondly, by using items as the units of the to-be-assessed events more comprehensive memory measures could be investigated (e.g., recall output position as a measure of ease of retrieval; see Chapter 2).

In all the trials, items were presented visually (details below). However, for some of these trials, participants were required to read the items out loud as they appeared. Reading presented items aloud has been shown to produce a modality effect (e.g., Conrad & Hull, 1968; Crowder, 1970); in other words, recency effects are usually enhanced for lists read aloud (as compared to lists read silently) in the same manner as for aurally presented lists (as compared to visually presented lists).

By manipulating reading instructions – i.e., either to read out loud or silently – our intent was to influence the accessibility of the last portion of the lists: Negative items presented in last positions which are read aloud were predicted to be better remembered than negative words presented in last position, but read silently. If a memory-based approach to RE (e.g., the Availability Heuristic; Tversky & Kahneman, 1973) is tenable, then it is predicted that lists which include negative items presented in last position and that are read aloud (experimental condition) will be rated as more unpleasant than lists where the negative items presented in last position and read silently (control condition).

3.2.1. Method

3.2.1.1. Participants

A total of thirty-three undergraduate students (30 females) from City University, London, took part in the experiment and were granted course credits for an introductory course in psychology. Age of participants ranged from 18 to 25 years (M = 19.4, SD = 2.1).

3.2.1.2. Design and Materials

A pool of 240 words was selected from the Affective Norms of English Words database (ANEW; Bradley & Lang, 1999). Thirty-six negative items were selected along with 204 neutral ones. As in Chapter 2, items were selected on the basis of their valence (i.e., their rating along the positive-negative scale) and arousal values (i.e., how "activated" the normative sample felt while reading the words).

Negative items were low in valence (less than 3 on a scale of 1-9) and high in arousal (greater than 5 on a scale of 1-9); neutral items scored in the middle range for valence (between 4.5 and 6.9) and relatively low on the arousal scale (less than 6). Selected negative items were overall lower on the valence scale (M = 2.2 and M = 5.7, respectively), $t(238) = 30.6$, $d = 6.3$; at the same time, negative items were significantly higher on arousal ratings (M = 6.3 and M = 4.2, respectively), $t(238) = 18.8$, $d = 4.5$.

Through within-list matching between negative and neutral items, forty-eight 5-word lists were created: Overall there were 12 "P1" lists (i.e., lists were a negative item was presented in first position and it was followed by 4 neutral items), 12 "P3" lists, 12 "P5" lists and finally 12 "C" lists, which included only neutral words. Items were not repeated across trials.

In order to equate lists for their length – and for their frequency and familiarity ratings – they were assigned to four different groups of 12 lists each. The first three groups included "negative" lists (i.e., lists with 4 neutral and 1 negative items), while the 4th group included only "Control" lists (5 neutral items). The groups were equated on familiarity ratings (Coltheart, 1981) – as groups averages ranged from 521.4 to 539.8 – number of phonemes (from 4.4 to 4.8) and the Kucera-Francis frequency index (Kucera & Francis, 1967), with groups averages ranging from 36.4 to 51.0.

Negative items were rotated, so that each of the 36 negative items selected from the database (ANEW; Bradley & Lang, 1999) appeared in each possible position (1st, 3rd and 5th) an equal number of times across participants.

Finally, block rotation across participants allowed each of the 48 lists to be read out loud and silently an equal number of times; also, each list was presented as often in the first half of the experiment (24 lists) or in the second half (24 lists). This is because the reading instructions were counterbalanced across participants: Half of the time participants performed silent encoding in the first half of the experiment, while the other half of the time participants were told to read out loud the first 24 lists of the experiment.

3.2.1.3. Procedure

Participants were tested individually in a session that lasted approximately 50 minutes. The experiment was run on a program developed using Authorware 7.0 (Adobe / Macromedia, 1987, 2003). In the first set of screens, participants provided demographic data and were given the chance to become accustomed to the computer-controlled procedure.

Participants were told that the aim of the study was to collect normative data about the pleasantness of word lists. They were instructed to attend to the lists and to provide an overall pleasantness rating for each list after its presentation. The ratings were collected on a 0-100 scale (0 = *very unpleasant*, 100 = *very pleasant*), and participants were encouraged to make use of the whole range in their responses when using the slide bar and its marker (see Experiments 2 and 3).

A memory task followed the rating task: As in Experiment 3, participants were told to type in the first two words that came to mind from the just-attended list, and then to click the "Continue" button. In the next screen, participants had the opportunity to type in any additional words they might remember, however it was stressed that it was not necessary to type in any further word and that it was fine to proceed directly to the next trial. Instructions also stressed that the rating and the memory task were equally important and instructed participants not to neglect the former in order to proceed more quickly to the latter.

As mentioned above, each experimental session was divided into two halves of 24 lists each, separated by a short break which lasted no more than 5 minutes. The only difference between the 2 halves was the reading condition (silent or aloud).

Finally, three practice trials allowed participants to become familiar with the whole procedure. List presentation order was randomised for each participant and both the rating and memory tasks were self-paced.

3.2.2. Results

As in Chapter 2, there was a predetermined criterion for a participant's data to be included in the analyses: any participant who produced judgment scores with a standard deviation lower than 5 (5% of the scale) was excluded. No participants were excluded from the analyses as the lowest standard deviation for the judgment scores was 8.8. Alpha was set to .05 for all analyses.

Memory. The memory manipulation was fairly successful and the modality effect was observed regardless of item valence (see Figures 3.1a and 3.1b below). Overall, items read out loud (M = .72, SD = .12) were better remembered than items read silently (M = .67, SD = .10). However, the memory advantage for aloud items was only due to the last position (M = .89 and M = .71, respectively). For the remaining two positions, memory performance was rather comparable between the two modalities.

As expected, negative items (M = .79, SD = .13) were better remembered than neutral items (M = .67, SD = .12). Finally, overall serial position effects were observed, as items presented in first (M = .75, SD = .18) and last positions (M = .80, SD = .10) were better remembered than words presented in the middle position (M = .64, SD = .14).

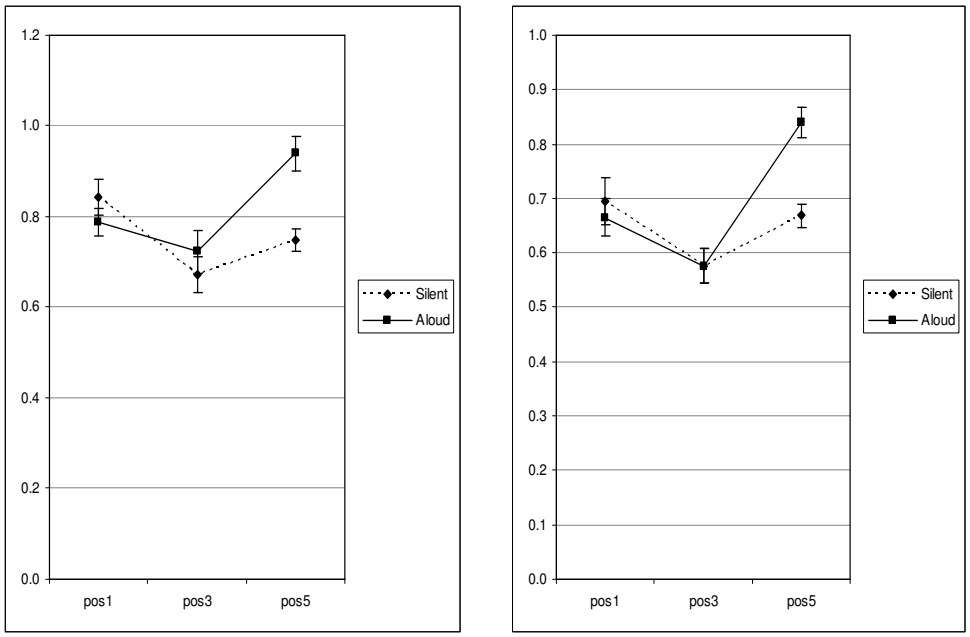


Figure 3.1 Mean proportion recall as a function of Modality (Silent vs. Aloud) and Position (1, 3, and 5) – for negative (Fig.3.1a, left) and neutral (Fig.3.1b, right) items. Error bars indicate *SEM* (Exp.4)

The memory performance results were analysed with a 3-way repeated measures ANOVA with Position (positions 1, 3 and 5), Valence (Negative vs. Neutral) and Modality (Silent vs. Aloud) as factors. The analyses revealed that the only significant interaction was between Position and Modality, $F(2, 64) = 15.9$, $p < .001$, $\eta_p^2 = .33$, which confirmed that the recall advantage of aloud items compared to silent items was observed only in the last (recency) position ($p < .001$).

The significant main effect of Valence $F(1, 32) = 16.7$, $p < .001$, $\eta_p^2 = .34$, confirmed that negative items were better remembered than neutral ones – and this was true regardless of Position and Modality (both Fs < 1) . The main effect of Position $F(2, 64) = 18.5$, $p < .001$, $\eta_p^2 = .37$ and the contrast analyses revealed how items presented in positions 1 and 5 were better remembered than

items presented in 3rd position (both *p*s < .001) – but did not differ between each other (*p* > .11).

To summarise, the memory manipulation was successful: the modality effect was observed together with primacy and recency effects. If participants accessed information in memory in order to provide summary assessments, then it would be predicted that lists which were read out loud and which included a negative item in last position should be rated more unpleasantly than lists where the recency negative item was read silently. This is because recency items which were read aloud were more accessible in memory – a fact that is confirmed by a higher recall rate for those items as compared to recency items read silently.

Moreover, as overall, negative items in primacy and recency positions were better recalled than negative items presented in the middle position, it was predicted that primacy and recency effects should be observed in retrospective judgment, too. Lists with a negative item in either the first or last position should be rated as more unpleasant than lists where the negative item was presented in third position – and this should be true regardless of modality.

Judgment. As in Chapter 2, the pleasantness ratings were transformed as follows: the average pleasantness ratings for the Control lists for each participant were subtracted from the pleasantness ratings for each P1, P3 and P5 lists. These corrections were made according to modality: the average ratings of Control lists read aloud were subtracted from the ratings of P1, P3 and P5 lists which were also read aloud, and the same was done for lists read silently. Therefore, the corrected judgment scores (J') indicated how much more unpleasant each list was rated as compared to the Control lists. The correction was performed in

order to reduce the extent of anchoring affects (e.g., Chapman & Johnson, 2002) and the inter-individual differences in the use of the rating scale. The following analyses were run on the J' scores[9].

Overall, neither the presentation position of the negative items nor modality affected pleasantness ratings (see Table 3.1 below). Pleasantness ratings was relatively equivalent across the board – despite an overall tendency for lists read aloud (M = -13.0, SD = 9.9) to be rated as more unpleasant than lists read silently (M = -11.2, SD = 9.7). Also, lists P3 lists were rated as slightly more pleasant (M = -11.1, SD = 10.3) than P5 (M = -12.2, SD = 9.3) and P1 lists (M = -13.1, SD = 9.5).

Table 3.1 Mean corrected pleasantness ratings (J') as a function of list type and modality (Exp.4)

		List type		
		P1	P3	P5
Modality				
Silent	*M*	-12.5	-9.6	-11.6
	SD	(13.1)	(11.7)	(12.1)
Aloud	*M*	-13.7	-12.6	-12.8
	SD	(10.6)	(12.8)	(12.8)

A 2 × 3 repeated-measures ANOVA was run with List type (P1, P3 and P5) and Modality (Silent vs. Aloud) as the independent variables. None of the

[9] The same analyses were performed also on the raw scores, which revealed that 1) the Control lists were rated as significantly more pleasant than all the other lists types (all ps < .001) and 2) the retrospective judgment pattern as a function of list type ratings was virtually the same as for the J' scores.

effects were significant (all *p*s > .31) – confirming that pleasantness ratings were low regardless of modality and of the presentation position of the negative item.

3.2.3. Discussion

The results of the present experiment seem to contradict the hypothesis that memory moderates RE: manipulating memory availability for distinctive items within the lists did not lead to the predicted changes in pleasantness evaluations. The modality effect (e.g., Corballis, 1966; Murray, 1966) was observed for the lists recall: negative items presented in last position and read out loud were better remembered than the corresponding negative items read silently. If participants relied on the items accessible in memory in order to assess the lists pleasantness, then it seems reasonable to expect that lists where the negative item in the last position was read out loud should have been rated as more unpleasant than corresponding lists which were encoded silently. Instead, virtually no difference was observed in the retrospective judgment scores for these two list types.

More strikingly, compared to Experiments 1, 2, and 3, neither primacy nor recency effects were observed in RE: that is, contrary to previously replicated findings (see Chapter 2), the presentation position of a negative item inserted in a list of neutral items did not differentially affect the pleasantness ratings for the list as a whole.

On the basis of these results it could be argued that there is no functional relationship between memory and judgment processes – any possible association between the two cognitive functions arising as "spurious" effects (e.g., due to some stimuli dimensions affecting both in the same way; Anderson, 1989). Along the same line of reasoning, it could be argued that primacy and recency

effects observed in memory and judgment in Experiments 2 and 3 (see Chapter 2) might have been due to separate causal dynamics (see Hogarth & Einhorn, 1992; Tan & Ward, 2000). This however, can be seen as giving a null finding significant theoretical weight.

One hypothesis about the findings of Experiment 4 is related to modality related interference effects (e.g., Glenberg, 1984). Perhaps performing a visual evaluation task interfered with the recall for the recency items more when the said items were presented in the visual modality – as compared to when they were read aloud. In other words, the visual distractor task (pleasantness ratings) would involve more item overwriting with visual recency items (read silently) than with auditory recency items (read out loud). According to this hypothesis, when participants attended silently to the lists, the negative item presented in last position affected their judgment; however, providing the overall assessment would have interfered with the same modality list representation, reducing memory for the recency items. The extent of this judgment-driven *interference* effect on recall for recency items would have been lower for items which were read aloud –because the modality of their presentation was different from the modality in which the evaluation task was performed (e.g., Nairne, 1990). It could be argued that the observed memory advantage for recency aloud items over the recency silent items was probably much smaller at the time of judgment, as it was artificially inflated by the differential distractor effects of the RE task over silent and aloud lists. This could explain why pleasantness ratings for "P5" lists were comparable regardless of the modality of presentation.

In order to assess the reliability of these findings, an additional experiment was run. The general objectives remained the same; namely, to significantly

affect pleasantness ratings of lists by manipulating the memorability of the distinctive items within the lists. At the same time though, significant methodological changes were implemented. In order to limit any potential confounding effects of recency overwriting, the memory manipulation involved items presented in the *middle* positions of the lists.

Secondly, in order to reduce the accuracy of the list representation accessed in memory, list length was increased to 7 items. This modification will be coupled by manipulating memorability of items in more than one presentation position – which will also allow us to increase the number of critical trials which will be analysed.

Lastly, the memory manipulation which will be implemented will allow us to be more confident that the representation of the lists accessed in memory at the time of judgment will be closer to that accessed at the time the memory task is prompted.

Introductory note

The two following experiments have been submitted as a paper for publication, which is currently awaiting the editor's decision.

3.3. Abstract

The nature of the memory processes involved in retrospective evaluations remains an object of debate in the literature. Some theoretical approaches (e.g., "two-memory" hypothesis; Anderson, 1996) propose functional independence between memory for and judgment of verbal stimuli: summary evaluations are based upon the on-line integration of moment-by-moment evaluations made during the actual event. On the other hand, several empirical findings suggest that judgment depends on how easily information can be retrieved from memory (e.g., Schwarz, 1995). In the present study, the nature of the relationships between memory and judgment was investigated. In Experiment 5, participants recalled word lists after rating each list for pleasantness. The effect of a negative item on retrospective evaluations was moderated by its accessibility, which was manipulated. The possibility that these results were due to the dual nature of the task was ruled out in Experiment 6 – where participants were presented with a surprise memory task. The memory-judgment relationship was maintained with this incidental learning paradigm. The results support a memory-based approach to retrospective judgment and suggest that people retrieve episodic information about events in order to evaluate them in hindsight.

3.4. Introduction

In everyday life, it is a regular experience to evaluate events once they have unfolded. Retrospective evaluations (hereafter RE) are coherent evaluations which involve the integration of information from hedonic states into a unitary judgment. Summary assessments can be provided about episodes which may have varied in quality and intensity over time (Fredrickson, 2000) and about target stimuli which have been presented in a sequential manner (e.g., Lichtenstein & Srull, 1987). Examples of the type of experiences which have been considered in the literature range from continuous annoying sounds (Schreiber & Kahneman, 2000) to the enjoyment experienced during a recent vacation (Kemp, Christopher, & Furneaux, 2008). Evaluations of sequential information include personality impression formation from lists of attributes (e.g., Anderson & Hubert, 1963) and Eurovision contestant evaluations (Bruine de Bruin, 2005).

An important question concerns the nature of the memory processes involved in RE: Do people retrieve episodic information about experiences in order to evaluate them in hindsight? In the literature, there are two contrasting approaches to this issue that are more widely documented. On one hand, some theoretical accounts (e.g. "the two-memory" hypothesis; Anderson, 1996; Anderson & Hubert, 1963; the "value-account"; Betsch, Plessner, Schwieren, & Gütig, 2001) propose functional autonomy between memory and judgment processes. According to these accounts, impressions of episodes are formed "on-line" (i.e., while they are being experienced) and RE are constructed (cf. Fredrickson, 2000) from the product of this on-line judgment. Retrieving episodic information from the event itself is viewed as a cognitively costly

operation – which is called upon if, and only if, the on-line judgment is prevented (for example, by not forewarning participants about the subsequent judgment task; Hastie & Park, 1986). These approaches propose that people do not rely in any significant manner on episodic information about an event when they evaluate it in hindsight.

On the other hand, several theoretical views suggest that – to some degree or another – retrieved information influences judgment often leading to biases in RE (e.g., Dougherty, Gettys, & Ogden, 1999; Tversky & Kahneman, 1973; see also Schwarz, 1995). For these approaches, the moments within an episode that are most available in memory ("accessible", cf. Schwarz, 1995) disproportionately affect retrospective judgment. In support of the role of retrieval in summary assessments, many studies have found significant correlations between memory and judgment measures, suggesting that memory and judgment may be functionally related (e.g., Aldrovandi, Poirier, & Ayton 2008; Moser, 1992; Reyes, Thompson & Bower, 1980; Schwarz, 1995; Tversky & Kahneman, 1973).

However, correlations between memory and judgment do not preclude the possibility that such correlations are attributable to other factors (e.g., vividness; Shedler & Manis, 1986); it remains possible then that retrospective judgment is not causally related to retrieval processes. For instance, Anderson (1989) suggested that *impression memory* (i.e., on-line judgment) and *verbal memory* (i.e., episodic memory) may be "distinct functionally" (p. 209) but the output from the two systems may correlate since they operate on the same attended stimuli (see also Moser, 1992). To re-iterate, the suggestions is that even if 'memory for' and 'judgment of' verbal stimuli are significantly related this

does not necessarily imply that people base their retrospective judgments on the episodic information they retrieve from memory.

The aim of the present investigation was to implement a stricter test of the hypothesis that people access information about a specific event in order to evaluate it. Somewhat stronger evidence in support of the role of memory in RE would be produced if manipulating the memorability of certain moments within the to-be-assessed episodes influences retrospective judgment. If RE do not depend on the information retrieved from memory, then manipulating the accessibility of some segments within the events should not affect judgment in hindsight. On the other hand, if retrieval and judgment processes are functionally dependent, then the easier it is to access specific information, the larger its impact on RE.

To our knowledge, few studies so far have attempted to predict judgment as a consequence of experimental memory manipulations (e.g., Gabrielcick & Fazio, 1984; Hanita, Gavanski, & Fazio, 1997; Lewandowsky & Smith, 1983). Lewandowsky and Smith (1983) increased the memorability of non-famous instances within a set through repetition, which in turn increased the corresponding frequency estimates participants provided. The authors concluded that the successful memory manipulation affected the participants' judgment responses (see also Gabrielcick & Fazio, 1984).

However, participants in the above mentioned study were asked to complete an evaluation task that involved frequency estimation of specific instances within a search set – a typical example of frequency judgment (e.g., Tversky & Kahneman, 1973). There is very little evidence that bears upon the role of memory in retrospective evaluations which in all likelihood do not rely on

the same cognitive processes as frequency estimation (see Hogarth and Einhorn, 1992, for a discussion of this issue).

Aldrovandi et al. (2008) showed that inserting a negative item within a list of neutral words significantly reduced the judged pleasantness of the list – relative to lists that did not contain a negative item. In the present paper, we called upon the same judgment task and manipulated the memorability of these negative items. The rationale was that if a negative item's accessibility in memory is heightened, then its impact on the summary assessment of the list as a whole should be larger than if the negative item's accessibility is not increased.

3.5. Experiment 5

In order to further investigate the role of memory in RE, word-lists were used as the to-be-assessed events. The major reasons behind this choice were that 1) mnemonic behaviour for this type of material is well documented and 2) it was possible to control the characteristics of the material that are thought to affect memory and judgment. For example, we could control for familiarity, frequency, stimuli length, valence, and arousal value of the items called upon. These factors are known to affect retrieval from memory and as such –if they are not controlled— may also influence the memory judgment relationship in a number of spurious, undesirable ways. Moreover, using words allowed us to use more comprehensive memory measures, including item *accessibility* which was measured through recall output position (see Aldrovandi et al., 2008). Briefly, we assumed that items recalled first were more accessible in memory than items recalled later or not at all.

The experimental manipulation involved increasing the accessibility of a negative item within some of the lists – and partially hindering it within others. Item accessibility was boosted by increasing both temporal (e.g., Brown, Neath & Chater, 2007) and perceptual isolation (e.g., Hunt, 1995; Rabinowitz & Andrews, 1973). The latter was achieved by presenting target negative items in a red font – while all the other items were presented in a regular black font. At the same time, the isolated items were also preceded and followed by a temporal gap, contrary to the other items. These manipulations were chosen as it can be argued that they would directly impact memory without affecting the pleasantness of the words themselves.

It was predicted that isolated negative items would be better remembered than non-isolated negative words. If episodic memory is accessed when people provide summary assessments about an episode, then the lists where the negative item was isolated should be rated as more unpleasant than the lists where the negative item was not isolated.

3.5.1. Method

3.5.1.1. Participants

A sample of thirty-two undergraduate and postgraduate students and staff members (23 females) from City University London took part in this study. Age ranged from 18 to 54 (M = 27.3, SD = 8.1). Participants were recruited through e-mail advertising and were either granted course credits for an introductory

course in psychology or were awarded with £8 in exchange for their participation.

3.5.1.2. Design and Materials

A total of 294 words were selected from the Affective Norms of English Words database (ANEW; Bradley & Lang, 1999). Thirty-two items were selected as negative items and 262 were chosen as the neutral items. Words were selected depending on their scores along the database dimensions of valence (how positive or negative they are) and arousal (how much self-appraised activation the normative sample reported while reading the word; Bradley & Lang, 1999). Negative items were low valence (less than 3 on a scale of 1-9, 1 being the most negative) and relatively high in arousal (higher than 5, on a scale of 1-9). Neutral items scored in the middle range for valence (between 4 and 7) and relatively low on the arousal scale (lower than 5). The selected negative items were significantly lower than neutrals on valence scores (M = 2.2 and M = 5.8, respectively; $t(292) = 27.1$, $d = 3.2$) and higher on arousal scores (M = 6.3 and M = 4.3, respectively; $t(292) = 16.3$, $d = 2.1$).

Forty-two seven-item lists were created in the following manner: 32 "Negative" lists included six neutral and 1 negative item, while the remaining 10 lists were "Neutral" lists, as they contained only neutral words. The negative lists differed depending on the position of the negative item: Overall there were two P1 (i.e. lists where the negative item was presented in first position), P2, P6 and P7 lists and eight P3, P4 and P5 lists. It was decided to consider the P1-P2 and P6-P7 list types as "fillers" in order not to rely on data gained from primacy and recency positions as these are more prone to ceiling effects; given our main

manipulation involves increasing the memorability of items that are already relatively distinct, primacy and recency positions were not adequate for our purposes.

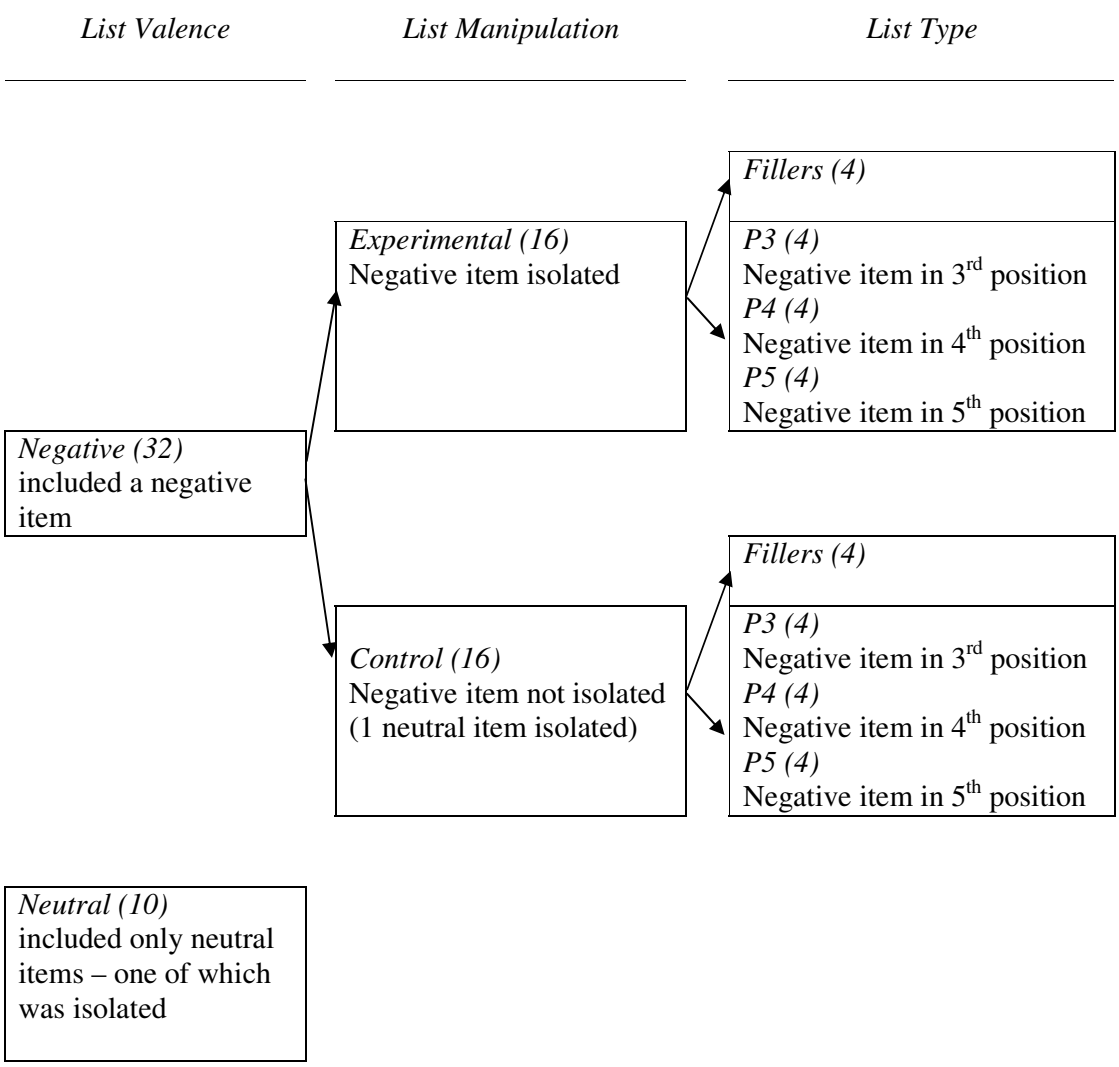
Within-list matching between the negative and the neutral items was implemented in order to ensure that negative and neutral words were equated on familiarity ratings (Coltheart, 1981), number of phonemes, and the Kucera-Francis frequency index (Kucera & Francis, 1967). Items were not repeated across trials.

In order to manipulate memory accessibility, two well established and reliable memory effects were called upon and implemented simultaneously. First, item accessibility was increased by temporal isolation (e.g., Brown et al., 2007). Despite the debate about the role of time in short time memory (e.g., Lewandowsky, Duncan, & Brown, 2004), there is clear agreement regarding the empirical effect that is obtained when an item is preceded and followed by a 0.5-sec interval; it will be better recalled than a corresponding non-isolated item. The second manipulation involved making the temporally isolated item physically more distinctive (e.g., Rabinowitz & Andrews, 1973) – by presenting it in red while all the other items were presented in black.

To sum up, in every list, there was a temporally isolated, red, item. Within each experimental list containing a negative item, the said item was presented in red and it was temporally isolated. On the other hand, there were an equal number of control lists where the negative item was neither temporally nor physically isolated – but a neutral item was. The isolated neutral item was presented 1 or 2 positions away from the non-isolated negative item – an equal number of times in each direction. For instance, for each participant, one P3

control list had a neutral isolated item in position 1, one in position 2, one in position 4 and the last in position 5. Please see Table 3.2 below for a summary of the list types utilised in the present experiment.

Table 3.2 Lists utilised in Exp.5 (number of lists in brackets)



The decision to implement two memory manipulations simultaneously originated from a pilot study (n = 7), where it was ascertained that manipulating the physical distinctiveness of the items alone did not produce a sizeable memory advantage. In order to achieve greater confidence in any results, it is argued that

a medium to large memory advantage is necessary; such a change in memorability should involve a corresponding change in RE. In this context it is important to bear in mind that the RE concerns the list as a whole, while the memory manipulation only touches upon a single item.

Neutral lists also had 1 of the neutral words presented in red and temporally isolated. Overall, the number of times a red item – irrespective of its valence – appeared in each of the positions was equated as much as possible across list types; a red item appeared in each of the 7 positions with a frequency ranging from 5 to 7.

Finally, the 32 negative lists were divided in 16 pairs of matched lists (along the dimensions of familiarity, number of phonemes and frequency). Rotation across participants allowed each of the lists to be presented the same number of times as experimental and as a control lists – and the same number of times as a P3, P4, P5 or filler list.

3.5.1.3. Procedure

All participants were tested in an individual session that lasted approximately 45 minutes. The whole experimental procedure was controlled by a computer program that was developed using Authorware 7.0 (Macromedia; 1987, 2003).

A series of screens gathered demographic data and allowed the participants to familiarise themselves with the computer-controlled procedure.

Participants were told to attend to the word lists, presented on the screen, one at a time. They were told that a randomly selected item within each list was going to be presented in red – and were also instructed not to pay particular

attention to it. After the last presented word disappeared, a series of asterisks were displayed for 3 seconds, indicating the end of the list; participants were then to complete the rating task, by providing an overall pleasantness rating for the list they had just attended to. Ratings were collected on a 0-100 scale (0 = *very unpleasant*, 100 = *very pleasant*), and participants were encouraged to use the whole range of scores. In order to provide the ratings, participants used the mouse to click along a slide bar (with extremes of 0 and 100) on the position they felt best reflected their impression of the list. In order to reduce the possible influence of anchoring effects (e.g., Chapman & Johnson, 2002) a slider marker appeared on the bar only after participants clicked on the slide bar for the first time. Participants could then adjust the initial rating and confirm their final assessment by clicking on a "Continue" button.

Immediately after having provided the assessment, participants were required to perform an easy memory task. Participants had to type in the first two items that came to mind from the just-attended list. It was stressed how spelling mistakes would not affect scoring. Participants then clicked on the "Continue" button to access the following screen, where they had the chance to type in any other word they might remember. However, it was stressed in the instructions how not remembering any other item apart from the initial 2 was fine. The present memory task was utilised so as to not distract participants from the judgment task (see Aldrovandi et al., 2008). Moreover, the instructions stressed how both the rating and recall tasks were equally important and encouraged participants not to overlook the former in order to proceed more quickly to the latter.

Both the rating and recall tasks were self-paced and list presentation was randomised independently for each participant. Also, participants had the chance to become familiar with the task, through 3 practice trials. A post-experimental questionnaire was used at the end of the session in order to gather information about how participants completed both the judgment and recall tasks.

As a check on the effectiveness of the memory manipulation, 14 participants were also asked to perform a further task, during which they had to perform only a memory task after each list was presented. The 21 lists used for this second half of the experiment were a random sub-sample of the 42 lists used for the experiment. Participants were told to recall the first two words that came to mind, and to type in any other word they might remember once the first two answers were provided. Results showed that the negative items (M = .67, SD = .17) were better remembered than neutral items (M = .55, SD = .17), $t(13) = 2.9$, $p < .05$, d = .78. Overall, isolated items (M = .68, SD = .16) were better remembered than non-isolated items (M = .54, SD = .18), $t(13) = 2.6$, $p < .05$, d = .69, hence suggesting that the memory manipulation was effective.

Importantly, as Valence and Isolation increased memory performance when participants only had to perform a memory task, any memory advantage observed in Experiment 5 cannot be due simply to on-line judgment increasing items memorability (e.g., Alba & Hasher, 1983; Hastie, 1981).

3.5.2. Results

Lists with a Peak either in 1st, 2nd, 6th or 7th were excluded from the analyses – as they were fillers. Alpha was set to .05 for all analyses.

Memory. Overall, the memory manipulation appeared successful: Isolated negative items (*M* = .70, *SD* = .26) were better remembered than non-isolated negative items (*M* = .53, *SD* = .29) – and this was true across positions. A slightly less sizeable but still consistent memory advantage was also observed for the isolated neutrals (*M* = .53, *SD* = .31) when compared to non-isolated neutrals (*M* = .43, *SD* = .19). Regardless of Isolation, negative items (*M* = .62, *SD* = .16) were overall better remembered than neutral items (*M* = .48, *SD* = .15), and this pattern again seemed true across positions. Finally, memory for the items in 5th position (*M* = .64, *SD* = .27) was slightly better than memory for the items presented in 3rd and 4th positions (*M* = .56, *SD* = .27 and *M* = .57, *SD* = .25, respectively). See Figure 3.2 below.

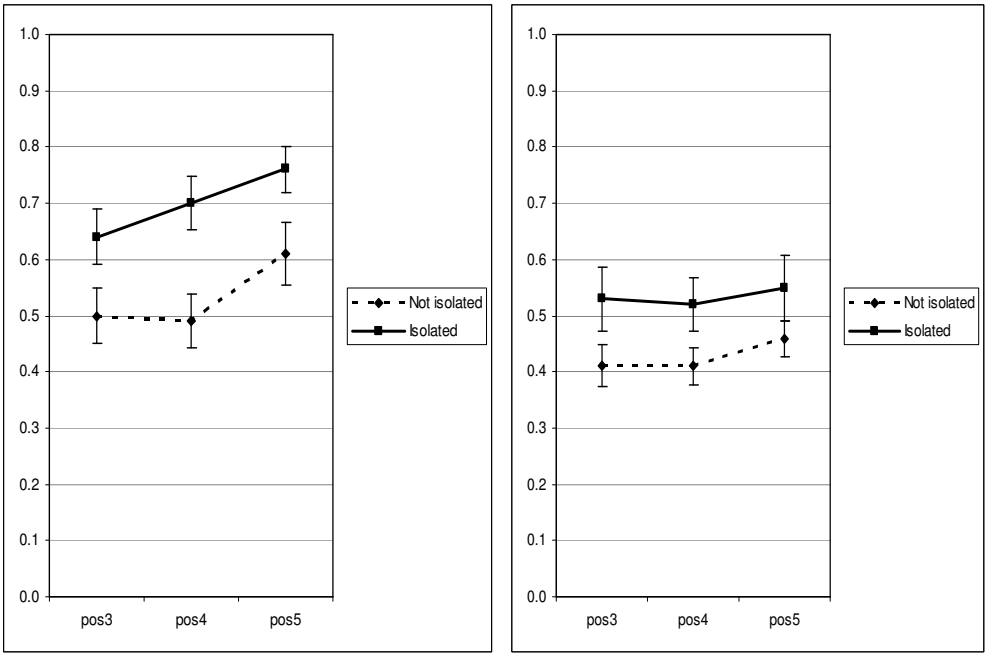


Figure 3.2 Mean recall proportion as a function of word position and isolation for negative (Fig.3.2a, left) and neutral (Fig.3.2b, left) items. Error bars indicate *SEM* (Exp.5)

A 3-way repeated measures ANOVA was run, with Position (positions 3, 4 and 5), Valence (Negative vs. Neutral) and Isolation (Isolated vs. Not-isolated) as factors. The analyses revealed that none of the interactions were significant. The significant main effect of Valence, $F(1, 31) = 22.1$, $p < .001$, $\eta_p^2 = .42$, confirmed that overall negative items were better recalled than neutrals. The lack of a significant interaction between Valence and Position, $F(2, 62) = 1.1$, $p > .32$, indicated that the memory advantage was constant across positions. The significant main effect of Isolation, $F(1, 31) = 13.4$, $p < .001$, $\eta_p^2 = .30$, confirmed that isolated items were better remembered than not-isolated ones – and this was again true regardless of Position ($F < 1$) and Valence ($F(1, 31) = 2.9$, $p > .09$). Finally, the main effect of Position, $F(2, 62) = 4.1$, $p < .05$, $\eta_p^2 = .12$, together with the post-hoc analyses revealed how items presented in 5th position were better recalled than items in 3rd and 4th positions. It can be concluded that the memory manipulation was successful, as isolated negative items were better remembered than non-isolated negative words. Moreover, the isolation effect applied to neutral items as well.

If a memory-based approach to RE is tenable, then it would be predicted that the experimental lists (i.e. those lists where the negative item was isolated) should be rated as more unpleasant than control lists.

Judgment. In order to limit the extent of possible anchoring effects (e.g., Chapman & Johnson, 2002) and to reduce the influence of inter-individual differences in the use of the 0-100 scale, judgment scores were transformed as follows (see Aldrovandi et al., 2008). The average pleasantness rating for the Neutral lists for each participant was subtracted from the pleasantness ratings for

each P3, P4, and P5 lists[10]. The new corrected judgment scores (J') therefore represented how much more unpleasant each P3, P4 and P5 list were in comparison to the average Neutral list for each participant (see Table 3.3).

Table 3.3 Mean corrected pleasantness ratings (and SD) as a function of isolation and list type (Exp.5)

Isolation/List Type	*P3*	*P4*	*P5*	*Total*
Negative item isolated	-16.5 (15.0)	-19.9 (16.6)	-22.1 (15.7)	-19.5 (15.7)
Negative item not isolated	-11.2 (10.9)	-10.8 (8.4)	-14.7 (13.6)	-12.2 (10.7)
Total	-13.9 (12.7)	-15.3 (12.4)	-18.4 (14.7)	

It is immediately apparent that, overall, lists with an isolated negative word (M = -19.5, SD = 15.7) were rated as more unpleasant than lists with a non-isolated negative word (M = -12.2, SD = 10.7). Also, this pattern is true irrespective of presentation position. Finally, ratings for lists with a negative item in 5th position (P5; M = -18.4, SD = 14.7) were slightly lower than the pleasantness ratings for P3 and P4 lists (M = -13.9, SD = 12.7 and M = -15.3, SD = 12.4, respectively).

A 2-way repeated measure ANOVA was run with List Type (P3, P4 and P5) and Isolation (Isolated vs. Not-isolated) as factors. The analyses yielded main effects of Isolation, $F(1, 31) = 24.2$, $p < .001$, $\eta_p^2 = .44$, and of List type, $F(2, 62) = 3.4$, $p < .05$, $\eta_p^2 = .10$. Post-hoc analyses revealed how pleasantness

[10] Analyses performed on the raw data revealed the main effect of List type (P3, P4, P5 and Neutral) was highly significant, $F(3, 93) = 35.9$, $p < .001$. Post-hoc testing confirmed that the Neutral lists were rated on average as more pleasant than any list containing a negative item (all ps < .001)

ratings for P5 lists were lower than for P3 lists ($p < .05$), but no differences were noted between P4 lists and the remaining two list types ($ps > .10$). The interaction between List type and Isolation was not significant ($F < 1$), indicating that lists with an isolated negative item were rated as more unpleasant than lists with a corresponding non-isolated item, regardless of presentation position.

The pattern of pleasantness ratings described above is completely compatible with the predictions based on the memory results: negative items presented in red and temporally isolated were better remembered, so lists including this type of word were rated as more unpleasant. This pattern was observed across presentation positions. Moreover, it can also be noticed that the memory advantage for the Peak in 5th position was reflected in more unpleasant judgments for the lists with a Peak in 5th position.

Memory-Judgment relationships. As for the analyses on Judgment only, the analyses of the relationship between memory and judgment were performed on corrected judgment scores.

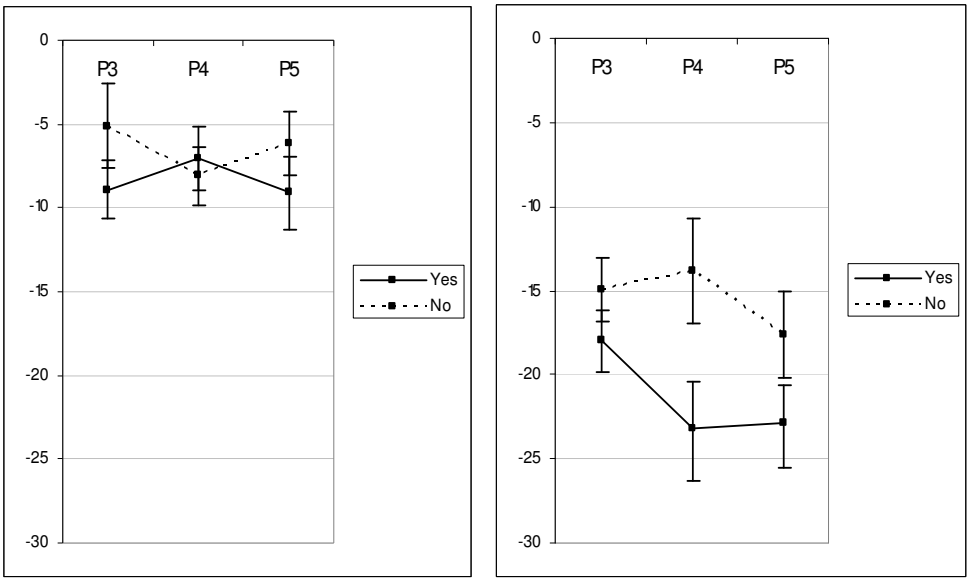


Figure 3.3 Mean corrected judgment (J') as function of isolation (Yes vs. No) and list type (P3, P4, and P5) – when the "Peak" was not recalled (Fig.3.3a, left) and when it was recalled. Error bars indicate *SEM* (Fig.3.3b, right; Exp.5)

Figures 3.3a and 3.3b illustrate how memory and judgment were clearly associated. As can be seen, overall, ratings in Figures 3.3a are higher than ratings in Figure 3.3b. The lists for which the negative item was not recalled shown in Figure 3.3a (M = -7.5, SD = 12.0) were rated as considerably more pleasant than the lists for which the negative item was recalled (M = -18.5, SD = 16.1) – which are summarised in Figure 3.3b.

Furthermore, it seems that Isolation had an adverse effect on judgment only for those lists where the negative word was recalled (see Figure 3.3b): In this case, Isolated lists (M = -21.5, SD = 13.1) were rated as more unpleasant than not-isolated lists (M = -15.4, SD = 12.4). However, when the negative item was not recalled (see Figure 3.3a), Isolated lists (M = -8.2, SD = 8.3) were rated similarly to non-Isolated lists (M = -7.4, SD = 5.4).

A 2 (Isolation: Yes Vs. No) × 2 (Memory: negative item recalled Vs. not recalled) × 3 (List type: P3, P4 and P5) repeated measures ANOVA was run on

the data summarised in Figures 3.3a and 3.3b[11]. The analyses revealed a main effect of Memory, $F(1, 31) = 28.8$, $p < .001$, $\eta_p^2 = .48$ – confirming that, overall, lists where the negative word was recalled were rated significantly lower than lists where it was not recalled. The main effect of Isolation was also significant, $F(1, 31) = 8.5$, $p < .01$, $\eta_p^2 = .21$. Overall, lists with the isolated negative word ($M = -14.9$, $SD = 14.2$) were rated as more unpleasant than lists where it was not isolated ($M = -11.1$, $SD = 13.9$).

The only significant interaction was between Memory and Isolation, $F(1, 31) = 6.6$, $p < .05$, $\eta_p^2 = .18$, which revealed how Isolation affected judgment only when the isolated negative item was recalled. On the one hand, lists whose isolated negative item was recalled were rated as more unpleasant than lists whose non isolated negative item was recalled, $t(31) = 4.0$, $p < .001$, $d = .71$. On the other hand, judgments were not significantly different between control and experimental lists when the negative item was not recalled, $t(31) = 0.9$, $p > .38$.

An important point to discuss in the above findings relates to one of the general patterns of data: lists with isolated negative items were associated with more unpleasant ratings only when the negative item was recalled. One interpretation of this finding invokes accessibility (e.g., Schwarz, 1995) – i.e. the ease with which the target negative item was retrieved from memory. A memory-based approach to judgment suggests that item accessibility moderates summary assessments, as information that is brought to mind most easily affect judgment most heavily.

[11] This analysis yielded 12.8% of missing data, which were missing completely at random as the MCR Little's test was not significant, χ^2 (156) = 159.5, $p > .40$. Missing data were replaced via mean by subject substitution.

In order to test this idea, recall or output position was used as a measure of accessibility. The pleasantness ratings were averaged according to the position in which the negative item was recalled by the participants. The underlying rationale was that items that are more easily accessible in memory are more likely to be recalled as earlier responses. Thus, if the negative item is more accessible and it is recalled early, it is argued that its impact on judgment would be higher than when it is less easily accessible.

Table 3.4 Mean corrected judgment (and SD) as a function of "Peak" recall output position (Exp.5)

		Negative item recall output position		
		Recalled in pos 1	Recalled in pos 2	Recalled in pos 3-7
Pleasantness ratings	*M*	-25.3	-18.4	-13.4
	SD	(15.7)	(18.0)	(9.0)

Overall, it seems clear that the pleasantness ratings were related to the output position of the negative item: They were lowest when the participants recalled it as the first response (M = -25.3, SD = 15.7) and they were highest when it was recalled amongst the last 5 outputs (M = -13.4, SD = 9.0). When the negative item was recalled as the second response (M = -18.4, SD = 18.0), the pleasantness ratings were at an intermediate level (see Table 3.4).

A significant one-way ANOVA[12], with negative item output position as the factor (position 1, position 2 and positions 3 to 7), confirmed these observations, $F(2, 62) = 12.8$, $p < .001$, $\eta_p^2 = .29$. Post-hoc analyses showed that

[12] This analysis had 2.1% of missing data, which were missing completely at random as the MCR Little's test was not significant, χ^2 (1) = 2.1, $p > .15$. Missing data were imputed via mean by subject substitution.

when the negative word was recalled as the first response, pleasantness ratings were lower than when it was recalled as the second response, $t(31) = 3.0$, $p < .01$, $d = .53$. Also, the difference in judgments between lists whose negative item was recalled as second response and lists where it was recalled among the last five outputs was significant, $t(31) = 2.1$, $p < .05$, $d = .37$. These analyses clearly suggest that an item's accessibility moderates its impact on retrospective judgment.

Next, we asked whether the memory manipulation increased accessibility of a negative item (e.g., the ease with which it was retrieved from memory) – beyond its availability (e.g., its recall rate). This could explain why, in those situations where the negative item was recalled, experimental lists were rated as more unpleasant than control lists (see Figure 3.3b). If the isolated negative items were more easily accessible in memory, their effect on pleasantness ratings should be larger than non-isolated negative words. Obviously, when the negative item was not recalled, accessibility did not moderate RE (see Figure 3.3a); it follows that when the negative item was not reported as a response during the memory task it was less likely to be accessed in memory at all.

For each participant we computed the proportion of trials where the negative item, when recalled, was either the 1st, 2nd or 3rd to 7th response. Table 3.5 below summarises the findings.

Table 3.5 Mean proportion recall (and SD) for negative items, depending on their isolation and recall output position (Exp.5)

	Negative item recall output position		
Negative item Isolation	Recalled in pos 1	Recalled in pos 2	Recalled in pos 3-7
Isolated	.42 (.24)	.21 (.16)	.37 (.27)
Not Isolated	.31 (.29)	.19 (.16)	.50 (.32)

Table 3.5 suggests that when a negative word was isolated it was more readily retrieved from memory, as it tended to be recalled as an early response more often than non isolated negative items. When isolated negative items were recalled, they were more likely to be recalled as the first response (on average 42% of the times) than the non isolated ones (31%); at the same time, non isolated negative items (50%) were more likely to be recalled amongst the last five responses than isolated ones (37%).

An ANOVA was run, with the proportion as the dependent variable and Isolation (yes vs. no) and Recall output position (1st vs. 3rd to 7th)[13] as the factors. The analyses revealed no main effects of Isolation and Recall output position

[13] The 2nd recall output position was excluded from the analyses in order to avoid violating the independence assumption of ANOVA.

(both Fs < 1). More interestingly, the interaction between Isolation and Recall output position was significant, $F(1, 31) = 9.9$, $p < .01$, $\eta_p^2 = .24$. Follow-up analyses confirmed that isolated negative items were recalled more often as first response than non isolated negative items, $t(31) = 2.7$, $p < .05$, $d = .48$ – while the opposite was true for the 5 last output positions, $t(31) = 2.8$, $p < .01$, $d = .55$.

These results suggest that isolated negative items, when recalled, were more easily accessible than non isolated ones. This in turn can explain why, when the negative item was recalled, those lists including an isolated negative word were rated as more unpleasant than lists where the negative item was not isolated.

To summarise, detailed patterns of pleasantness ratings were successfully predicted by the memory performance participants exhibited. Negative items were better remembered when isolated than when non-isolated. As expected if RE are based, at least in part, on the information available in memory, pleasantness ratings including the former type of negative items were lower than for lists which contained the latter type.

Furthermore, the accessibility of negative items was increased by the experimental manipulation, too. When recalled, the isolated negative words were more easily accessible in memory than non-isolated negative items – as they tended to be recalled more often as early outputs. Increased negative item accessibility was associated with a larger impact on RE.

We would submit that the pattern of results described above provides compelling evidence in support of a memory-based approach to RE. However, one argument that has recurred when discussing this work needs to be addressed: the effect of the dual nature of the task. When performing the evaluation task,

participants knew that they had to perform a memory task subsequently. It could be argued that the nature of the experimental task artificially induced participants to rehearse 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 on the same lists artificially created the patterns obtained in Experiment 5. Experiment 6 addressed this issue.

3.6. Experiment 6

In order to test our interpretation of the results of Experiment 5, an additional experiment was run. Participants were presented with a *surprise* memory task. As the participants could not anticipate this memory task, there is no reason for the memory component of the procedure to have any impact on 1) the retrospective judgements and 2) the association between what is recalled and the pleasantness ratings.

In the present experiment, participants were presented with an evaluation task for a number of trials – only to be asked to perform a *surprise* memory task after the very last pleasantness rating was provided. If the findings of the previous experiment are reproduced when participants are not aware of the memory task, then the suggestion that the paradigm used in Experiment 5 artificially increased the associations between memory and judgment can be ruled out.

3.6.1. Method

3.6.1.1. Participants

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

3.6.1.2. Design and Materials

A sub-sample of 20 lists was selected from those used for Experiment 5, 15 of which included a negative item and six neutral items (Negative lists) and five which were "Neutral" lists, as they contained 7 neutral items. The negative item lists were divided into three blocks of five lists each. The blocks were closely matched on familiarity ratings (blocks averages ranging from 521.6 to 534.4), number of phonemes (from 4.0 to 4.8) and the Kucera-Francis frequency index (59.2 to 68.5).

The five negative items within each block were matched with the negative item within the other two blocks, still on the dimension of familiarity (517.2 to 533.4) number of phonemes (4.6 to 5.2) and frequency (36.6 to 56.4). Negative items were rotated, so that each of the 15 negative items appeared in each possible position (1st, 4th and 7th) an equal number of times across participants.

Each participant was presented with four trials in total; in the first trial they were always presented with an all-neutral item list randomly selected from

the block of Neutral lists. The remaining three trials included one each of the three negative list types, i.e. "P1" (where the negative item was presented in first position), "P4" and "P7". A Latin Square design allowed us to present the different list types the same number of times as the last trial across participants – the last trial was the critical trial as it was followed by the surprise memory task.

3.6.1.3. Procedure

The present experiment was presented after another task, during which participants had to assess short slideshows – there was no memory component in this other task. Prior to any testing, participants were presented with a series of introductory screens which provided general information about the study and gathered demographic data.

Participants were told that they were to assess the pleasantness of word lists. When participants had rated the 4th and last list, a screen instructed them to recall the first two words that came to mind from the just-presented list; if they could not remember any word they just had to press the "Continue" button twice. After participants entered their responses, they had a chance to enter any other word they might have remembered from the list in a following screen.

3.6.2. Results

Information about the IP address and the time participants took to complete the whole experimental procedure was collected. No data were gathered from the same IP address, so no data were excluded on the basis of this criterion. Six participants (i.e., 5.3% of the total sample) were excluded from the

analysis because of excessive task duration, as it took them from 51 to 203 minutes to complete the whole experiment (while the average duration was 17 minutes and the maximum duration allowed was 30 minutes.)

Availability. As for Experiment 5, the analyses were run on corrected judgment scores (J'): these scores represented how more unpleasant the negative list presented in the last trial was rated as compared to the Neutral list[14].

An independent samples t-test was run, with J' as the dependent variable and Memory (negative item recalled vs. not recalled) as the between-subjects factor. The t-test was significant, $t(105) = 10.6$, $p < .05$, $d = .52$, and revealed that the pleasantness ratings for those lists where the negative item was not recalled ($M = 0.3$, $SD = 20.7$) were significantly higher than for those lists where the negative item was recalled ($M = -10.3$, $SD = 19.9$).

Accessibility. Next, the influence of negative item *accessibility* on pleasantness ratings was analysed. Pleasantness ratings were averaged as a function of negative item output position (see Table 3.6 below).

The pleasantness ratings were highest when the negative item was not recalled ($M = 0.3$, $SD = 20.7$) and lowest when it was recalled as the first response ($M = -13.0$, $SD = 21.3$). When the negative item was recalled as either 2nd, 3rd, or 4th item, the pleasantness ratings were at an intermediate level ($M = -3.7$, $SD = 14.7$).

[14] Analyses performed on the raw data revealed that the main effect of List Type (P1, P4, P7, and Neutral) on pleasantness ratings was significant, $F(3, 318) = 12.6$, $p < .001$. Post-hoc analyses confirmed that the Neutral lists were rated as significantly more pleasant than all the list types which included a negative item (all ps $< .05$).

Table 3.6 Mean corrected judgment (and SD) as a function of recall output position (Exp.6)

		Negative item recall output position		
		Not recalled (n = 38)	Recalled in pos1 (n = 49)	Recalled in pos 2-4 (n = 20)
Pleasantness ratings	*M*	0.3	-13.0	-3.7
	SD	(20.7)	(21.3)	(14.7)

A one-way between-subjects ANOVA was run, with negative item recall position (not recalled; recalled as 1st response; recalled as 2nd to 4th response) as the factor. The analysis yielded a significant main effect, $F(2, 104) = 4.9$, $p < .01$, $\eta_p^2 = .09$. Planned contrasts revealed that when the negative item was recalled as the 1st response, pleasantness ratings were significantly lower than when it was either not recalled or recalled as a later response ($p < .05$)[15]. At the same time, lists pleasantness when the negative item was recalled as 2nd to 4th response did not differ from when it was not recalled at all ($p > .47$).

Overall, the correspondence between memory and judgment was clear: pleasantness ratings were lower when the negative item was recalled and they were the lowest when it was most easily accessible in memory.

As this pattern was observed even if participants did not expect a memory task, it can be concluded that the paradigm called upon in Experiment 5 cannot be the source of memory-related judgment effects.

[15] For the latter comparison it was a 1-tailed test.

3.6.3. Discussion

The results of the two experiments reported here provide evidence supporting the hypothesis that memory plays a significant role in summary assessments. Retrospective judgment was successfully predicted on the basis of the memory pattern observed for the negative item (i.e., a distinctive episode within the event).

First, increasing the distinctiveness – and consequently the recall rate – of a negative item within a list of words led to the whole list being rated as more unpleasant, relative to situations where memorability was not enhanced.

Secondly, results suggest that the *ease of recall* (i.e., accessibility; Schwarz & Vaughn, 2002; Tversky & Kahneman, 1973) also moderates RE: the easier it was to access a negative item in memory, the larger its impact on summary assessments. When recalled, isolated negative items were more easily *accessible* than non-isolated negative items – as they were recalled more often amongst early responses. This resulted in pleasantness ratings for experimental lists (which included an isolated negative item) being lower than for control lists – obviously only when the negative information was more likely to be accessed in memory at the time of judgment (as it was recalled at the memory stage).

To summarise, the manipulations enhanced memory performance for the negative items by making them more often present in memory and, at the same time, by increasing the ease with which they were retrieved from the degraded trace of the just-seen list (see also Aldrovandi et al., 2008).

These results were not just a spurious effect of the isolation manipulation; neutral items were also better remembered when they were isolated, but this had no effect on judgment – something to be expected as *neutral* items would not

push RE in any specific direction. Also, the results were not due to the nature of the paradigm, which, it could be argued, might lead participants to access information in memory at the time of judgment simply because they knew they had to perform a recall task subsequently. The results of Experiment 6 clearly showed that memory-judgment correspondence was observed even when participants could not anticipate the memory task.

In summary, a memory-based approach to RE can account for the overall pattern of judgments and the details of the judgment behaviour participants exhibited. Other approaches (e.g., Anderson, 1989) do not fair as well, and could not have led to the predictions corroborated here. It can be argued that isolating the negative item did not *directly* affect the pleasantness of the lists – as the latter was unaffected by the experimental manipulation when the negative items were not recalled. Indeed, experimental and control lists were rated as equally unpleasant on the occasions in which the negative item was not recalled. This finding also ruled out the possibility that participants' overt strategies to comply with the experimental manipulation significantly affected the results. Indeed, it could have been argued that manipulating memorability in such an evident way (isolated items were both physically *and* temporally isolated) could have induced participants to guess the aim of the study and to engage in artificial overt strategies.

Finally, approaches that contend that it is instead judgment that causes memory – by biasing it at either the encoding or retrieval stages (Alba & Hasher, 1983; Hastie, 1981) – struggle to account for the present findings. These theories hold that the relationships between memory and judgment are a consequence of the direct influence of on-line evaluative processes on memory; to frame it in the

present paradigm, the argument would be that the increased memorability of isolated items was due to the decreased pleasantness of lists of words. However, this view does not seem tenable for at least two reasons. First, as previously mentioned, negative item isolation was associated with lower pleasantness ratings *only* when the negative item was recalled. Secondly, in the second part of the Experiment 5 (n = 14) it was shown that the memory advantage for isolated negative words was evident in the absence of a judgment task – hence without participants being prompted to form an on-line evaluation of the lists of words.

In conclusion, by implementing a strict test of the role of memory processes in the impression formation of events, it is argued that a stronger case can be put forward for the role of memory processes in RE. Summary assessments were significantly influenced by retrieval processes in the experiments reported here and accessibility moderated the impact of information on retrospective judgment.

Chapter 4:

Retrospective Evaluations of Stories

4.1. Introduction

The results of the previous two studies verified some claims concerning the role memory may play in biasing hindsight judgment. Chapter 2 consistently showed primacy and recency effects for summary assessments of word lists: a distinctive and valenced item inserted in a list of neutral words affected retrospective judgment to a larger extent if presented either at the beginning or at the end of the list (Experiments 1, 2 and 3). This finding was coupled with primacy and recency effects for memory; negative items presented either in first or last positions were better remembered than those presented in the middle of the lists (Experiments 2 and 3). Contrary to the predictions of "Independence" models (cf. Hastie & Park, 1986; p. 259), which postulate functional independence between memory and judgment processes (e.g. Anderson, 1989; Betsch, Plessner, Schwieren, & Gütig, 2001), correlational analyses revealed clearly that the two functions may be associated. First, availability as *content of recall* (hereafter "availability"; cf. Schwarz & Vaughn, 2002) moderated judgment – recalling a negative item was associated with lower pleasantness ratings. Second, availability as *ease of recall* (hereafter "accessibility"; Schwarz, 1995; Schwarz & Vaughn, 2002; Schwarz, Bless, Wanke, & Winkielman, 2003; Tversky & Kahneman, 1973) played a role, too: the easier it was to access in memory a negative item, the larger its effect on retrospective evaluations (Experiments 2 and 3).

Chapter 3 investigated further the associations between memory and judgment; the former was manipulated and the latter was successfully predicted. Enhancing memorability of a negative item resulted in decreased pleasantness ratings for those lists which contained it (Experiment 5). Moreover, participants'

awareness of the upcoming memory task did not bias the results: memory–judgment correlations were observed even when participants were asked to perform a surprise memory task (Experiment 6).

The aim of the experiments included in the present Chapter was to test the reliability of the above findings with to-be-assessed events other than lists of words. Can a memory-based approach predict retrospective judgments even when different types of stimuli (e.g., more ecologically valid) have to be evaluated? The question becomes most relevant as it could be argued that evaluating pleasantness of lists of words is a relatively abstract task, arguably seldom encountered in "everyday life".

More importantly, lists of semantically unrelated words represent discrete and rather non-cohesive events. It follows that, when assessing lists of words, it is unlikely that participants formed a 'gist' around the to-be-assessed and to-be-remembered events (Brainerd & Reyna, 1992). As it was difficult for participants to form a whole with the lists of words (they were semantically unrelated), it could be argued that they were "forced" to rely on the verbatim form of the inputs in order to perform both tasks – hence inflating the relationship between recall and judgment measures. In the literature, it has been shown that memory for the gist and for the verbatim materials show differential forgetting patterns, which argues for functional independence between the two types of representations (Brainerd & Reyna, 1992; Kintsch, Welsch, Schmalhofer, & Zimmy, 1990; Reyna & Kiernan, 1994). Further evidence from social cognition suggests that the degree of coherence of the to-be-assessed stimuli can moderate the relationship between memory and judgment (McConnell, Sherman, & Hamilton, 1994; 1997). McConnell and colleagues have argued that when

participants process stimuli which are expected to be low in *entitativity* (i.e., unity and coherence; cf. Campbell, 1958, cited in McConnell et al., 1997), they do not integrate information about the stimuli – which results in associations between memory and judgment. On the other hand, according to these authors, when stimuli are high in *entitativity* (e.g., "a person whose behaviour is quite predictable"; McConnell et al., 1997, p. 750), participants form an integrated evaluation of the stimuli, which leads to *on-line judgment* (see also Hastie & Park, 1986) and to independence between retrospective judgment and retrieval processes.

The main question that naturally arises is the following: "Will memory–judgment relationships be observed even if participants have the possibility of forming a more integrated representation of the to-be-assessed experiences?" Results from Moser (1992) suggest that different relationships between summary assessments and memory arise depending on the memory representations that are being tested (see also Experiment 2). More specifically, Moser (1992) found that memory and judgment did correlate when "self-generated" memory measures were used (i.e., participants were asked to describe the reasons behind a specific evaluation) – but no correlations were observed when standard recall measures were used (i.e., participants had to enumerate the information presented to them about a specific to-be-assessed scenario.)

In the present series of experiments, short stories told both through slideshows and spoken narratives will be used as the to-be-assessed events. It is argued that this choice of stimuli will increase the generalisability of the results described in the previous Chapters, which referred to lists of semantically unrelated words.

First, slideshow stories are, despite still being discrete in nature (as the slides will be presented sequentially), more cohesive and unitary events (i.e., higher in "integrity"; Loewenstein & Prelec, 1993; p. 93).

Second, they are arguably more ecologically valid stimuli. Evaluating short stories that are being told to us is a more frequent activity than rating the pleasantness of lists of words which are semantically unrelated.

Third, short stories told through slideshows should induce participants to form a *gist* of the to-be-assessed stimuli. This last point is possibly connected to the previous one, as it has been argued that with naturalistic stimuli (Weber, Goldstein, & Barlas, 1995; p. 54) participants are more prone to encode the gist of the stories in order to facilitate further decision-making (Brainerd & Reyna, 1992). Would this in turn lead participants not to retrieve specific information from memory in order to provide summary assessments (Brainerd & Reyna, 1990; 1993; Reyna & Brainerd, 1992)? Short slideshow stories, which in the present series of experiments typically involved the activities of a hypothetical character, should facilitate the creation of a more integrated or related representation of the to-be-assessed event (e.g., increasing their "*schematization"*; cf. Heuer & Reisberg, 1990; see also Bartlett, 1932; Minsky, 1975; Rubin, 1986).

In relation to the studies described in Chapters 2 and 3, this chapter investigates whether a relationship between memory and judgment can be obtained in situations where there are compelling reasons to expect an *on-line judgment* formation (Hastie & Park, 1986). Indeed, participants were told in advance about the evaluation task and the to-be-assessed stimuli should allow participants to integrate the information easily (as previously argued). If the

approaches which postulate independence between memory and judgment processes (e.g., Anderson, 1989; Betsch et al., 2001) are correct, then little or no associations between memory and judgment measures should be observed. On the other hand, if judgment depends, at least in part, on retrieving of information from memory (e.g., Dougherty, Gettys, & Ogden, 1999; Schwarz et al., 2003; Tversky & Kahneman, 1973), then correlations between memory and judgment measures should arise.

4.2. Experiment 7

The aim of the first experiment was to investigate whether summary assessments for slideshow stories produced a pattern similar to that observed for word lists. No memory task was used (see also Experiment 1, Chapter 2). However, in the second experiment of this series memory for the stories is assessed also; in this way, it will be possible to indirectly determine whether performing a memory task alters the retrospective judgment pattern.

The to-be-assessed events were short stories, told through a 6-slide presentation and spoken descriptions – the latter somewhat like a story being told. The choice of stimuli allowed us to both control variables that may affect judgment (i.e., valence and arousal) and to partial out the effects of factors that may influence memory processes (i.e., picture distinctiveness and familiarity, spoken descriptions length and grammatical structure).

Importantly, as for word lists, it was possible to manipulate within-subjects the point at which a negative event happened within the story. For instance, a story may have started with something negative happening, only to be

followed by relative neutral events and actions; or, the negative event might have happened in the middle of the story, or at the end. In this way, it was possible to investigate position effects on the summary assessments, i.e. primacy and recency effects.

4.2.1. Method

4.2.1.1. Participants

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

4.2.1.2. Design and Materials

A pool of 96 pictures was selected from the International Affective Pictures System (IAPS; Lang, Bradley, & Cuthbert, 1999). Twelve pictures were selected as negative slides, while 84 were chosen as neutral slides. Negative slides were low in valence (less than 3, on a scale ranging from 1 to 9, 9 being the most positive) while neutral slides scored in the middle range for valence (from 3.8 to 5.9 on the same scale). As a result, the former type of slides (M = 2.1, SD = 0.5) were significantly lower in valence than the latter (M = 4.5, SD = 0.7), $t(94)$ =

12.3, d = 4.0. The negative slides (M = 4.0, SD = 0.7) were also higher in arousal ratings than neutral slides (M = 2.4, SD = 0.6), $t(94) = 8.5$, d = 2.5.[16]

From the resulting slide pool, 16 six-slide stories were created, as follows. Four stories included a negative slide in the first position followed by five neutral slides – hereafter identified as "Start" (or "S") stories. Four "Middle" stories ("M") had a negative slide either in third or fourth position[17]. Four "End" stories ("E") encompassed five neutral slides and a negative one in last position. Finally, four "Control" ("C") stories contained only neutral slides. Slides were not repeated across trials.

Negative slides (M = 4.2, SD = 0.8) were generally lower in familiarity ratings (Libkuman, Otani, Kern, Viger, & Novak, 2007) than neutral slides (M = 5.8, SD = 1.1). Hence, it was decided to include a neutral picture within each story that included a negative picture so that the two were closely matched for familiarity. For example, within story 3, the negative slide depicted two elderly people and was characterised by a familiarity rating of 4.0 – while the matching neutral item showed a boy playing chess and had a familiarity rating of 4.1. Overall, the negative slides and the 12 matched neutral slides were very comparable in regard to familiarity ratings (both M = 4.2). Also, the presentation positions of the matched neutral slides mirrored those of the negative pictures, as follows: For the Start stories, the neutral slide was presented in last position, for the Middle stories it was presented in the middle position adjacent to the negative slide and for the End stories it was presented in first position. As negative and

[16] For arousal ratings there was overlap between negative and neutral slides: The former ranged from 3.2 to 5.5 (on the same scale from 1 to 9), while the latter ranged from 1.2 to 3.9. This could not be avoided as negative pictures which were very high in arousal usually involved scenes of revolting (e.g., a mutilated body) or threatening nature (e.g., an attacking snake)

[17] Contrary to the studies reported in Chapters 2 and 3, no distinctions will be made between P3 lists and P4 lists; this is because participants were only presented with two trials for each of these two list type

matched neutral slides were rotated across participants, any potential bias arising from serial position effects should be minimised[18].

A few further controls were implemented. The 16 stories were then assigned to four different groups of four stories each, with groups A, B and C as "negative" stories (which included five neutral and one negative slide) and group D consisting of the "Control" stories (six neutral slides). The four groups were closely matched on distinctiveness ratings (Lang et al., 1999), with group averages ranging from 3.0 to 3.6 (on a scale from 1 to 9), memorability (Lang et al., 1999; from 2.6 to 3.1), number of words (7.6 to 8.0) and phonemes (31.6 to 34.7) used for the narratives accompanying each slide.

As previously mentioned, negative items were rotated, so that each of the 12 negative words selected from the database appeared in each possible position (Start, Middle, and End) an equal number of times across participants. This allowed for each story to be presented in three different versions, across participants – without having the overall *gist* changed (see Appendix). For instance, story 8 described a part of a day in the life of *Sophie*, who *works as a legal secretary*. In the Start version of the story, *Sophie thinks of her sister who committed suicide* (the negative slide, depicting a cemetery), then *goes to the office downtown, on her way she sees a man playing music; at home, she looks out on the lake,* then *she takes a stool back to the kitchen.* In the Middle version of the story, we learn that Sophie is thinking of her sister after we find out she is a legal secretary, that she is in the office downtown and that she saw a man playing music – but before the two slides referring to the home environment (i.e., she is looking at the lake and she moves the stool). In the End version of the

[18] 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

story, we hear about who she is and her actions both at the office and in the house – only to learn at the end that she thinks of her sister. Three different versions were created for each of the four Control stories, too; this will allow us to check whether re-ordering the slides within a story (still without changing its overall *gist*) will affect retrospective judgments even if no negative event was included. Importantly, the wording which accompanied each slide did not change between the three different versions, so that no difference in either memory and/or judgment outputs could be attributed to the wording of narratives.

Each story was accompanied by a recorded narrative. Every effort was made to equate these recordings across stories. The same female voice was used for all the narratives. Also, the tone of the voice used was regular and did not express emotion. As mentioned before, within-story matching ensured that the number of phonemes and words used for each slide was comparable between negative and neutral slides. Moreover, the grammatical structure of all the sentences was comparable – they usually involved only a subject and a predicate. The recording that accompanied each slide did not exceed 3.1 seconds, so that some silent time preceded and followed each sentence (each slide was on the screen for a duration of 4 seconds.)

Two pilot studies were conducted: The first was run on the Web (n = 10) and the second one was run at the start of a lecture (n = 18). The following measures were taken based on participant feedback. First, some slide descriptions were changed so as to increase the correspondence between the pictorial information and the narrative – and to avoid "bizarreness" about the

unfolding of the events described[19]. Second, adverbs (e.g., "then") and conjunctions (e.g., "while") were dropped from the narratives – which made it possible to re-order the events of the stories without changing the wording. Lastly, the present tense only was used, in order to limit as much as possible inter-story differences.

4.2.1.3. Procedure

A series of screens gathered demographic data and allowed the participants to 1) familiarise themselves with the computer-controlled procedure 2) adjust the volume of their speakers/headsets and 3) accept the conditions described in the consent form. Participants were told the study was about their opinion of short stories told through slideshows – of how pleasant or unpleasant they were. They were instructed to attend to short slideshows, presented on the screen, one at the time. Each slide of the six-slide stories was displayed on the screen for a total duration of 4000ms, during which the corresponding narrative was played. No inter-stimuli interval was used – as the narratives themselves never lasted the whole 4 seconds (as previously mentioned, the longest one lasted for 3.1 seconds), hence allowing for some silent time at the beginning and at the end of each slide presentation.

After the last slide disappeared, a blank screen displayed for 500ms was followed by a series of asterisks displayed for 1000ms and another blank screen for 500ms. Participants then had to complete a rating task, by providing an overall pleasantness rating for the story they just attended to, *as a whole*. Ratings

[19] One picture (no. 2751) was also slightly altered during the process. A small portion of the picture was removed as it showed the profile of a man with a white beard – while the main character in the story so far was presented as a relatively young man.

were collected on a 0-100 scale (0 = *very unpleasant*, 100 = *very pleasant*), and participants were encouraged to use the whole range of scores. In order to provide the ratings, participants used the mouse to click along a slide bar (with extremes of 0 and 100) on the position they felt best reflected their impression of the story. In order to reduce the possible influence of anchoring effects (e.g., Chapman & Johnson, 2002) a slider marker appeared on the bar only after participants clicked on the slide bar for the first time. Participants could then adjust the initial rating and confirm their final assessment by clicking on a "Continue" button.

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 (see Appendix). Story presentation order was randomised independently for each participant and no time limits were set for the rating task.

4.2.2. Results and Discussion

Information about the IP address and the time participants took to complete the whole experimental procedure was collected. No data was gathered from the same IP address, so no data was excluded on the basis of this criterion. Six participants were excluded from the analysis because of excessive task duration, as it took them from 51 to 203 minutes to complete the whole experiment (while the average duration was 17 minutes and the maximum duration allowed was 30 minutes.) Three more participants were excluded as the standard deviation in their judgments was lower than 5 (i.e. 5% of the scale). This resulted in the total sample size being 104 (i.e. overall, 8% of participants were excluded.) Alpha was set to .05 for all analyses.

Figure 4.1 presents the mean pleasantness ratings as a function of story type. As expected, the Control stories (M = 65.9, SD = 18.2), with no negative slides/events, were rated as the most pleasant stories. Moreover, there is some evidence of recency effects in the ratings: End stories (M = 42.1, SD = 15.8)—which had a negative slide/event in the last position—received more negative ratings than the stories with a negative slide either in the first (M = 44.3, SD = 14.7) or in the middle positions (M = 44.7, SD = 14.3).

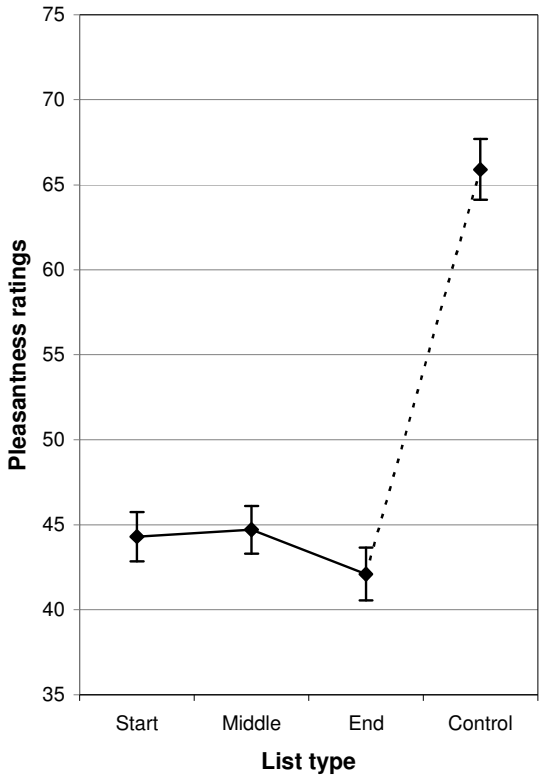


Figure 4.1 Mean pleasantness ratings as a function of story type. Error bars indicate *SEM* (Exp.7)

The inferential statistics confirmed these observations. A one-way ANOVA was run, with story type (Start, Middle, End, and Control) as the

within-subject factor. There was a significant main effect of story type, $F(3, 309) = 119.2$, $p < .001$, $\eta_p^2 = .54$. Planned comparisons revealed that pleasantness ratings for Control stories were significantly higher than for all the other story types (all ps $< .001$[20]). Judgments for End stories were lower than those for Middle ($t(103) = 2.2$, $p < .05$, $d = .22$) and Start stories ($t(103) = 2.0$, $p < .05$, $d = .19$); the ratings between the two latter story types did not differ between each other ($t(103) = 0.4$, ns).

Finally, a control analysis was run to check whether changing the order of the slides affected pleasantness ratings even for stories which did not include any negative slide/event (Control stories). A one-way ANOVA was run, with Version (1, 2, and 3) as the between-subjects factor, and confirmed that there was no difference between the three different versions of the Control stories, $F(2, 101) < 1$.

In sum, it can be concluded that the inclusion of a negative slide–accompanied by a narrative which describes a negative event–affected the perceived pleasantness for the story *as a whole*. Also, the negative effect of the negative slide was moderated by its presentation position: recency effects were observed for summary assessments as the negative event exerted a larger impact on pleasantness ratings when presented last – compared to when it was presented in comparison positions elsewhere in the stories.

When considering the judgment pattern, its differences with the judgments obtained for lists of words (e.g., Experiments 1, 2, and 3, Chapter 2) are immediately noticeable. Recency *and* primacy effects in retrospective judgment were obtained for lists of words, compared to only recency effects for

[20] For planned comparisons, significance levels will always refer to 1-tailed tests

slideshow stories; for the latter, it was noticed that End stories were rated as more unpleasant than Start stories.

It is not immediately clear why the judgment pattern differed depending on the to-be-assessed stimuli; however, addressing the following question may shed some light on the above discrepancy: "Can a memory-based approach explain the present findings?" If participants are accessing the representation of the just-seen story in order to evaluate it, then the results above suggest that the last events in the story weigh more heavily in the RE than the previous events. This would lead to the prediction that memory for the same events would be enhanced if a memory component was added to the task. Therefore, one interpretation of the reported results is that the recency effects in retrospective evaluations observed in the present Experiment are linked to a memory advantage for the negative slides which were presented in last positions. Experiment 8 examined this hypothesis.

4.3. Experiment 8

The main objective of Experiment 8 was to introduce a memory measure into the retrospective evaluation task in order to more directly investigate the relationship between retrospective judgment and memory for each story. As in Experiment 7, recency effects in retrospective judgment are predicted: stories that finish with a negative slide are expected to be more negatively assessed than the other stories – where the negative slide is presented either at the beginning or in the middle of the story. At the same time, it is expected that negative slides presented at the end of the stories will be characterised by a memory advantage

compared to negative (and distinctive) slides presented in the first and middle positions. Simply put, the prediction is that memory results will mirror the RE findings.

4.3.1. Method

4.3.1.1. Participants

A sample of 88 participants (55 males) took part in an internet-based experiment, advertised once again through ipoints®. Age ranged from 18 to 61 years (M = 42.3, SD = 10.2). Participants were granted ipoints in exchange for their participation.

4.3.1.2. Design and Materials

The materials employed were the same as in Experiment 7.

4.3.1.3. Procedure

As in Experiment 7, a series of screens gathered demographic data and allowed the participants to both familiarise themselves with the computer-controlled procedure and to adjust the volume of their speakers/headsets. Participants were again told that the aim of the experiment was to gather their opinion about the pleasantness of 16 short stories told through slideshows. After attending to a short slideshow and rating its overall pleasantness (on the same 0–100 scale), participants were asked to complete a memory task which involved two phases. During the first, participants were asked to type in a brief description

of the two pictures that came to mind most easily; participants were told that a simple description would suffice, and that spelling errors would not affect scoring. After participants entered the two descriptions, a new screen presented thumbnails of the six slides from the just-presented story, in random order. The participants' task was to select the two pictures that corresponded to the two descriptions they had just typed in – that is, they had to select the two pictures that came more readily to mind when thinking back at the last-presented story. The aim of the second part of this memory task was simply to validate participants' descriptions, and their selections will be used only to resolve any potential ambiguity that may have arisen from participants' descriptions entered in the first phase of the memory task.

As in Experiment 7, each slide was displayed for 4000ms – once again, with no inter-stimuli interval. Again, the last slide was followed by a blank screen for 500ms, a series of asterisks for 1000ms and another blank screen for 500ms. Then, the slider track appeared and participants performed the rating task, in a self-paced manner. The two memory tasks – both self-paced – followed the rating task, with no delay between them.

Participants were told that both the assessment and the memory tasks were equally important – they were asked not to overlook the rating task in order to proceed more quickly to the recall task. After participants had completed the recall task, the following list was presented, and so on.

Two practice trials were provided and included the same two stories used for the practice trials in Experiment 7. List presentation order was randomised independently for each participant.

4.3.2. Results and Discussion

No data were gathered from the same IP address, so this did not lead to any data exclusion. Two participants were excluded from the analyses as it took them 70 and 144 minutes to complete the experiment; the average duration was 24 minutes and the pre-determined maximum duration allowed was 45 minutes. Four participants were excluded because of the low standard deviation in their judgments (smaller than the pre-determined cut-off point of 5). Lastly, one participant was excluded from the analyses as s/he did not type the descriptions in the first phase of the memory test. The final number of participants was 81– that is, 8% of participants were excluded from the analyses. Alpha was set to .05 for all analyses.

Judgment. As in Experiment 7, recency effects only were observed for the pleasantness ratings. The Control stories (M = 64.6, SD = 17.6) were once again the most pleasant lists – as they included only relatively neutral slides/events. More importantly, End stories (M = 44.9, SD = 18.4) were rated as more unpleasant than Start (M = 48.1, SD = 17.4) and Middle stories (M = 47.4, SD = 17.5). Figure 4.2 below summarises these findings.

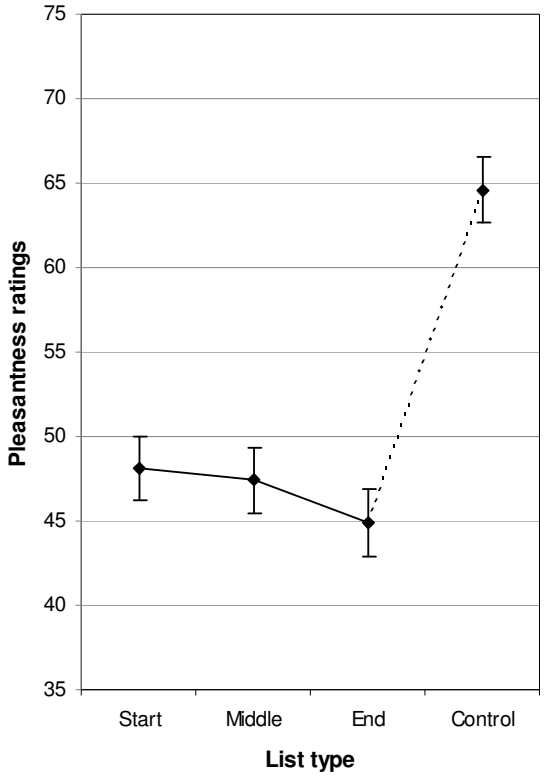


Figure 4.2 Mean pleasantness ratings as a function of story type. Error bars indicate *SEM* (Exp.8)

The significant one-way ANOVA – with story type (Start, Middle, End, and Control) as the within-subject factor – confirmed these findings, $F(3, 240) = 69.3$, $p < .001$, $\eta_p^2 = .46$. Control lists were again the most pleasant story type (all ps $< .001$). End stories were rated as more unpleasant than Middle ($t(80) = 2.1$, $p < .05$, $d = .23$) and Start stories ($t(80) = 2.9$, $p < .01$, $d = .32$). Start and Middle stories were rather comparable in regard to pleasantness ratings, $t(80) = 0.8$, ns.

As in Experiment 7, changing the order of the slides in the Control stories did not matter, $F(2, 78) < 1$.

Memory. As previously mentioned, recall for neutral slides will refer to the matched neutral items – equated for familiarity with the negative slides. Figure

4.3 represents the mean recall proportion as a function of slide position and valence. Overall there was a memory advantage for the negative slides as compared to the neutral ones presented in the same position; participants exhibited higher recall rates for the negative than for the neutral slides, across presentation positions. However, contrary to our predictions, *both* primacy *and* recency effects can be observed: Negative slides presented in either first (Start; $M = .63$, $SD = .29$) or last position (End; $M = .57$, $SD = .33$) were better recalled than negative slides presented in either 3rd or 4th positions (Middle; $M = .50$, $SD = .34$).

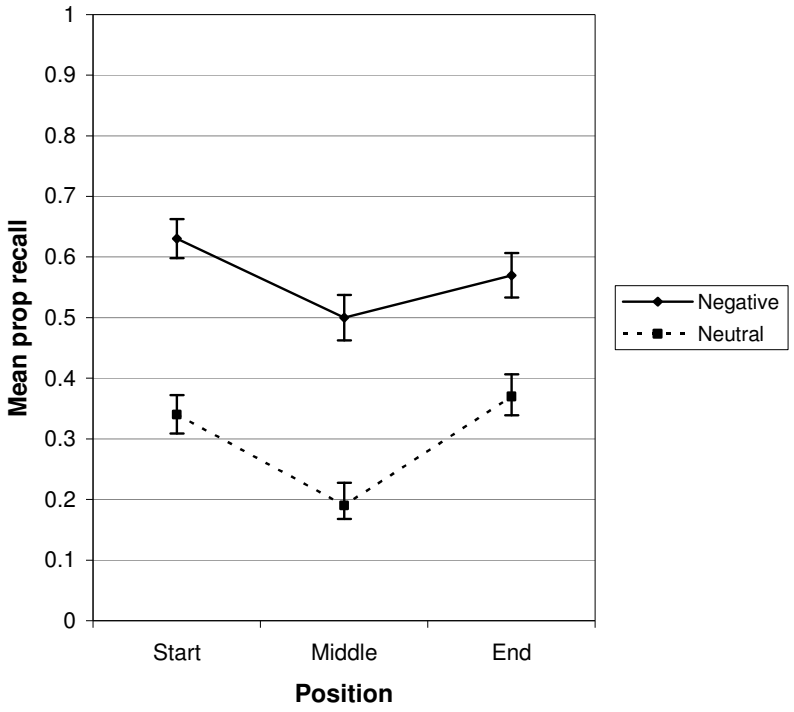


Figure 4.3 Mean recall proportion as a function of slide position and valence. Error bars indicate *SEM* (Exp.8)

The data were analysed using a 2 (valence: negative vs. neutral) × 3 (position: Start, Middle, and End) repeated measures ANOVA. Main effects of position ($F(2, 160) = 9.7$, $p < .001$, $\eta_p^2 = .11$) and valence ($F(1, 80) = 60.4$, $p < .001$, $\eta_p^2 = .43$) were noted. The slide valence by position interaction was significant, too ($F(2, 160) = 3.1$, $p < .05$, $\eta_p^2 = .04$); however, when correcting for the breached assumption of *sphericity*, the interaction became non-significant ($p > .09$).

Planned comparisons confirmed primacy and recency effects for the recall of the negative slides, as the recall for slides presented either in first and last positions was higher than recall for the slides presented in the middle positions ($t(80) = 2.8$, $p < .01$, $d = .33$; $t(80) = 1.9$, $p < .05$, $d = .22$, respectively). Recall for Start and End slides did not differ significantly, $t(80) = 1.3$, $p > .08$.

An overview of the results on pleasantness ratings indicates that they confirm the pattern observed in Experiment 7. Once again, a negative slide presented at the end of the story exerted the largest impact on summary assessments. However, the recall results did not support an approach where memory plays a central role in biasing retrospective judgments. Both recency *and* primacy effects were observed for the recall of the negative slides; slides presented in first or last positions were better remembered than slides presented in the middle of the story.

The discrepancy between memory and judgment patterns would suggest that the evaluations of the stories were independent of the memory representation that participants held about them. If participants relied at least in part on episodic information stored in memory when providing summary assessments, then it is reasonable to argue that Start lists, whose negative slide in first positions was

recalled to a higher extent, should have been rated as more unpleasant than Middle lists, where the negative slide was at a recall disadvantage.

A further finding goes against the predictions of a memory-based approach. If participants based their evaluations on the representations they stored in memory, *accessibility* (i.e., the ease with information is accessed in memory; e.g. Schwarz & Vaughn, 2002; Tversky & Kahneman, 1973) would have moderated retrospective judgment; in the present paradigm, it can be argued that situations where the negative slide was recalled as the first response (as opposed to second) should have been associated with the most unpleasant judgments. It also follows that negative slides presented in the last position should have been recalled most of the times as the first output, since the lists which encompassed them (the End stories) were rated as the most unpleasant. However, the recall output positions of the negative slides provided us with a rather different picture: Negative slides presented in last positions (End slides), when recalled, represented participants' first response only 42% of the times – compared to 63% for the Middle slides and 86% for the Start slides[21]. The earlier a negative slide was presented, the more likely it was recalled as first as compared to second response.

A compelling reason as to why the observed memory and judgment pattern were different may concern the nature of the representations accessed during both tasks – and the perhaps less than perfect correspondence between the two. The representation of the story accessed when evaluating it in hindsight might have been mostly *auditory* and semantic, as it concerned the *gist* of the

[21] Nevertheless, we are aware that participants were allowed to output two responses only and that this may have reduced the power of the recall output position as a measure of accessibility; after all, participants may have had vividly in mind more than two responses and simply chose two of them

story – its narrative and the events that shaped it. However, in the memory task participants were asked to recall the two *pictures* that came to mind most easily; hence the *visual* representation of the story elements was highlighted in the memory task. The two task instructions may have primed participants to access different information in memory about the same stimuli. After all, it can be reasonably assumed that the series of stimuli (the stories) was encoded in its multiple aspects, including 1) the pictorial information associated with the pictures 2) the gist, i.e. the result of integrative processes which combined into a unitary representation the different events and 3) the auditory information included in each slide description (the narratives were both played trough headphones/speakers and written below each slide). The two tasks may well have *cued* different information, which resulted in the dissociation between their outcomes.

In order to address the above mentioned issue, some changes were made in the procedure to reduce the discrepancy between what the summary assessments are based upon and what participants "use" when performing the memory task. These changes attempted to increase the correspondence between the nature of the information that is utilised when performing the pleasantness ratings and the trace of the just-experienced stimuli that is accessed in memory in order to comply with the recall instructions. Two parallel Experiments were run in order to achieve this goal in two different ways. In Experiment 9, the stimuli will be presented to participants in one modality only (visual) – to reduce the potential altering effects on the memory–judgment relationships due to different stimuli encoding modality. Participants in Experiment 10 will be presented with the *exact* same stimuli as in Experiment 8. However, instead of being asked to

recall the two *pictures* that come to mind most easily from the just-seen stories, they were asked to recall the two *moments* that were more readily available; this should allow for a better alignment between the memory and judgment tasks.

4.4. Experiment 9

In this Experiment, the effects of stimulus encoding modality on the correlations between memory and judgment will be investigated. In the previous Experiment, the stimuli were presented simultaneously in multiple modalities. First of all, the taped narratives were played through speakers/headphones (auditory). Second, a picture was presented on each slide (visual-pictorial). Lastly, the text of the narrative was displayed underneath the picture (visual-verbal). As previously argued, the rating and the recall in Experiment 8 may have rested on different dimensions of the encoded stimuli producing different patterns for retrospective judgment (recency effects only were observed) and memory performance (both primacy and recency effects took place.)

Therefore, it was decided to utilise only visual stimuli; a short narrative was displayed on each slide as printed text – neither taped descriptions nor pictures were presented to participants. It is argued that this measure should help to align the information accessed to complete both the RE and memory tasks. As a consequence, clearer associations between memory and judgment measures should be observed.

4.4.1. Method

4.4.1.1. Participants

A total of 81 participants (48 males) took part in an internet-based experiment, advertised through ipoints®. Participants' age ranged from 25 to 67 years (M = 44.1, SD = 11.3) and they were granted ipoints in exchange for their participation.

4.4.1.2. Design and Materials

The 16 stories used were the same as in Experiments 7 and 8. However, only the printed sentences referring to each of the six events (the narratives) were displayed on each slide.

4.4.1.3. Procedure

The procedure was the same as in Experiment 8. After having provided the pleasantness rating for each story as a whole, participants were asked to describe briefly the two moments that came to mind most readily. The following screen presented participants with the six printed sentences from the just-seen story, in scrambled order. Their task was to click on the two narratives that corresponded to the two descriptions they had provided in the previous screen; that is, they had to select the two narratives that came to mind most easily when thinking back at the story.

4.4.2. Results and Discussion

Once again, some participants were excluded from the analyses on the basis of pre-determined criteria. Four participants were excluded from the analyses as it took them from 130 and 4372 minutes to complete the experiment; the average duration was 23 minutes and the pre-determined maximum duration allowed was 45 minutes. Three participants were excluded because the standard deviation of their judgements was lower than 5. This time, no participants neglected the typing task, so no participant was excluded according to this criterion. The final sample size equalled to 74, which means that 8.6% of participants were excluded from the analyses. Alpha was set to .05 for all analyses.

Judgment. Pleasantness ratings for the different story types seem rather comparable; Start stories (M = 51.4, SD = 13.9) were rated slightly more pleasant than Middle (M = 49.3, SD = 13.7) and End stories (M = 48.5, SD = 13.1). Once again, the Control lists attracted the highest ratings (M = 69.0, SD = 15.9). Figure 4.4 below summarises these findings.

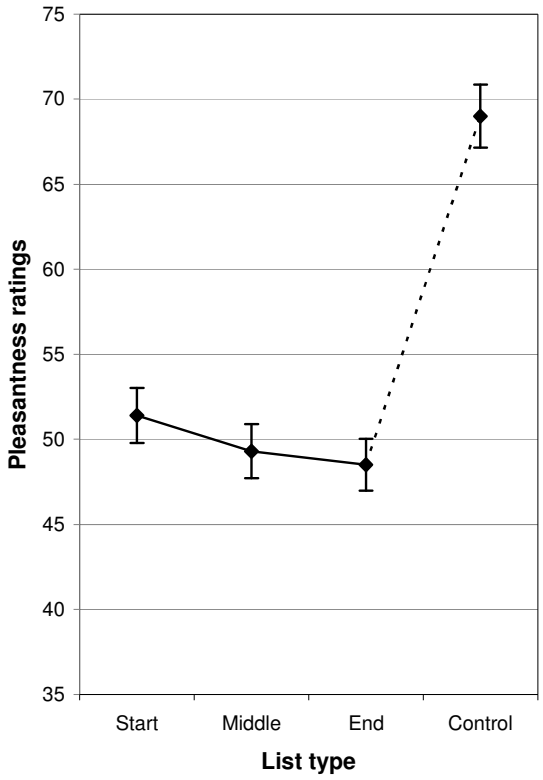


Figure 4.4 Mean pleasantness ratings as a function of story type. Error bars indicate *SEM* (Exp.9)

The main effect of story type was a significant, $F(3, 219) = 79.2$, $p < .001$, $\eta_p^2 = .52$. Planned comparisons once more revealed that the Control stories were rated as the most pleasant (all *p*s < .001). End lists were rated as more unpleasant than Start stories ($t(73) = 2.1$, $p < .05$, $d = .26$), while the difference between Middle and Start stories approached significance, $t(73) = 1.6$, $p = .06$.

As in the previous two Experiments, no significant difference were noted between the three different versions of the Control lists, $F(2, 71) < 1$.

Memory.[22] Figure 4.5 represents the mean recall proportion as a function of narrative position and valence. Negative narratives (M = .70, SD = .26) were better recalled than the neutral ones (M = .33, SD = .24) – and this memory advantage seems evident across presentation positions. Furthermore, recall for the negative narratives increased along with presentation positions, as narratives presented in last position (M = .74, SD = .26) were better recalled than those presented in the middle (M = .69, SD = .27) and first positions (M = .67, SD = .29).

A 2 × 3 repeated measures ANOVA was run, with Valence (negative vs. neutral) and Position (Start, Middle, and End) as the factors. The analysis revealed significant main effects of Position ($F(2, 146) = 14.2$, $p < .001$, $\eta_p^2 = .16$) and Valence ($F(1, 73) = 142.6$, $p < .001$, $\eta_p^2 = .66$). Additionally, the Valence by Position interaction was significant, too ($F(2, 146) = 4.1$, $p < .05$, $\eta_p^2 = .05$); follow-up analyses revealed that this was due to recency effects (i.e. the recall advantage of the narrative in last position compared to the recall for the narratives in middle positions) being larger for Neutral narratives ($t(73) = 4.9$, $p < .001$, $d = .57$), as compared to Negative narratives ($t(73) = 2.0$, $p < .05$, $d = .24$).

[22] As in the previous two experiments, the data for Neutral events will refer to the narratives matched for familiarity with the Negative events

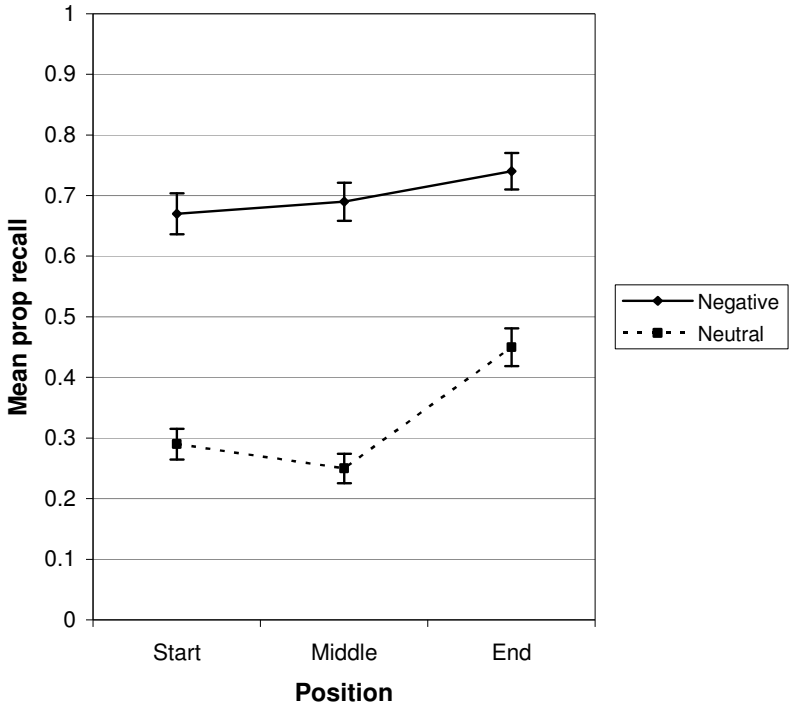


Figure 4.5 Mean recall proportion as a function of narrative position and valence. Error bars indicate *SEM* (Exp.9)

Finally, planned comparisons confirmed that recall for the End narratives (i.e. negative sentences presented in last position) was significantly higher than the recall of Start ($t(73) = 2.3$, $p < .05$, $d = .25$) and Middle narratives ($t(73) = 2.0$, $p < .05$, $d = .24$) – the latter two not being significantly different from each other, $t(73) = 0.5$, ns.

Memory-Judgment relationships. In order to reduce the influence of potential anchoring effects (e.g., Chapman & Johnson, 2002) and of inter-individual differences in the use of the 0-100 scale, judgment scores were transformed as follows: For each participant, the average pleasantness rating for the Control stories was subtracted from the pleasantness ratings for each Start, Middle, and

End story – that is the stories that contained a negative item. The new corrected judgment scores (J') therefore represented how much more unpleasant each Start, Middle, and End story was in comparison to the average Control story for each participant. J' scores were then averaged for each participant, according to the negative narrative presentation position and whether the negative narrative presented in the story was recalled or not[23,24].

Overall, when the negative sentence was recalled in the memory task, pleasantness ratings for the narrative as a whole were lower (M = -21.2, SD = 16.5) than when it was not recalled (M = -12.5, SD = 13.8). Moreover, this relationship between memory and judgment was relatively constant across story types. Table 4.1 below summarises these findings.

Table 4.1 Mean corrected pleasantness ratings (J') as a function of story type and the negative narrative being recalled or not (Exp.9)

		Story type		
		Start	Middle	End
Was the negative sentence recalled?				
No	*M*	-12.9	-17.8	-16.5
	SD	(15.7)	(14.1)	(15.8)
Yes	*M*	-20.8	-21.1	-21.8
	SD	(16.0)	(17.7)	(16.3)

[23] This analysis yielded a total of 29.7% missing values, which were missing at random as the MRC Little's Test was not significant, χ^2 (386) = 3.2, $p > .85$. 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, we will be reporting the data obtained via mean by subject substitution

[24] Due to the large amount of missing values (29.7%) correlational analyses were run between negative narrative recall proportion and average pleasantness ratings (averaging across trials). Analyses confirmed the significant associations between memory and judgment, in the expected direction (r = -.33, $p < .01$ for Start lists; r = -.45, $p < .001$ for Middle lists; r = -.29, $p < .01$ for End lists)

A 2 (Memory: Negative sentence recalled Vs. not recalled) × 3 (Story type: Start, Middle, and End) within-subjects ANOVA confirmed these observations. The main effect of Memory was significant ($F(1, 73) = 29.0$, $p < .001$, $\eta_p^2 = .28$), confirming that overall ratings were more unpleasant for those stories where the negative narrative was recalled. The relationship between memory and judgment was observed regardless of presentation position, as the Story type by Memory interaction was not significant ($F(1.6, 118.8) = 2.8$, $p > .09$)[25].

Finally, retrospective evaluations were analysed depending on the negative sentence recall position or output position (see Table 4.2); that is, the analysis was broken down into the situations when the sentence was recalled as the first response as compared to when it was recalled as second response[26]. The underlying rationale was that narratives that are more accessible in memory are likely to be recalled early, i.e. as the first response—if the negative sentence is more accessible and recalled early we would expect its impact on retrospective evaluations to be higher than when it is recalled as second response or not at all.

Overall, pleasantness ratings were slightly lower when participants recalled the negative sentence as the first response ($M = -22.6$, $SD = 18.1$) as compared to when they recalled as second response ($M = -20.6$, $SD = 15.2$). Judgments were most pleasant when the negative narrative was not recalled at all ($M = -12.5$, $SD = 13.8$).

The main effect of recall position on pleasantness ratings was significant, ($F(2, 146) = 23.0$, $p < .001$, $\eta_p^2 = .24$). Planned comparisons revealed that the

[25] Degrees of freedom were adjusted according to the Greenhouse-Geisser correction as the assumption of Sphericity was not met

[26] This analysis yielded a total of 6.4% missing values, which were once again imputed via mean by subject substitution

difference in judgments – between situations when the negative narrative was recalled as first as compared to second response – approached significance, $t(73) = 1.4$, $p = .07$. Hence, despite a trend in the expected direction, accessibility did not moderate retrospective evaluations.

Table 4.2 Mean corrected pleasantness ratings (J') as a function of negative narrative recall output position (Exp.9)

		Negative narrative recall output position		
		Not recalled	Recalled as first response	Recalled as second response
Pleasantness ratings (J')	*M*	-12.5	-22.6	-20.6
	SD	(13.8)	(18.1)	(15.2)

The pleasantness ratings results were comparable to those of Experiment 8 – as recency effects only were observed for summary assessments – however, the analyses of both memory measures and memory–judgment correlations revealed a rather different picture.

The recall results mirrored the pleasantness ratings. First of all, negative narratives were better remembered than corresponding neutral items, regardless of their presentation position within the stories. More importantly, the memory pattern for the negative narratives displayed recency effects only, as was predicted if the memorability of the narratives moderates the retrospective evaluations. This result provided preliminary support to the possible influence of memory processes on bias in judgment in hindsight.

Further analyses confirmed the associations between judgment and memory measures. First, pleasantness ratings were significantly lower for those stories where the negative narrative was recalled, as compared to when it was not

reported at all during the later memory stage. It is argued that when a negative narrative was not recalled in the memory task, it was also less likely to be available at the time of retrospective evaluations; on average, this would lead to a less negative assessment of the story.

Accessibility (as *ease of recall*; cf. Schwarz & Vaughn, 2002) was expected to moderate the effect of negative narratives on judgment; namely, it was predicted that the lowest ratings should have been provided when the negative narrative was recalled as the first response. Although the predicted pattern was not observed, there was a trend in the expected direction in the data.

In conclusion, by increasing the congruence in modality between the stimuli representations accessed during the judgment and memory tasks, clearer associations between memory and judgment pattern were observed. The next experiment attempted instead to increase the congruence between *what information* is used for judgment and memory tasks. Once again, based on the results of Experiment 8, recency effects only are predicted for memory and judgment patterns – together with clear associations between the two measures.

4.5. Experiment 10

In the present experiments, the stories were presented through both the visual and auditory modalities –as in Experiment 8– the narratives for each story were printed on the slides and played through speakers/headphones; moreover, they were accompanied by a picture. Participants were once again asked to rate the pleasantness of each story. However, one important change was implemented in regard to the memory task participants performed subsequently. Instead of

asking participants to describe the two *pictures* that come to mind most easily from the just-seen story, participants were asked to recall the two *moments* which are most readily available. The assumption was that this change would increase the correspondence between the information accessed by participants when performing the two tasks. It seems reasonable to argue that when participants are assessing the story for its pleasantness, they are thinking back at the different events which were described within the story. By asking participants to recall the most memorable moments, it is more likely that the same representation of the story is accessed in order to comply with the memory task instructions. This should in turn allow us to observe clear associations between memory and judgment, or at least *clearer* in comparison to Experiment 8, where participants accessed in memory the degraded pictorial trace of the story in order to perform the recall task.

4.5.1. Method

4.5.1.1. Participants

A total of 74 participants (36 males) took part in an internet-based experiment, advertised through ipoints®. Participants' age ranged from 23 to 68 years (M = 44.9, SD = 10.7) and they were granted ipoints in exchange for their participation.

4.5.1.2. Design and Materials

The materials employed were the same as in Experiments 7 and 8.

4.5.1.3. Procedure

The procedure was exactly the same as in Experiment 8. The only difference concerned the memory tasks. As soon as participants provided the summary assessment for a given story, participants were asked to "type a brief description of the two moments of the story that come to mind most easily". After participants entered the two descriptions, a screen displayed all the slides from the story, in random order. The participants' task was to click on the two pictures that corresponded to the two descriptions they had just provided[27].

4.5.2. Results and Discussion

Five participants (i.e. 7% of the total sample size) were excluded from the analyses according to the same pre-determined criteria as in the previous Experiments. Three participants were excluded from the analyses as it took them from 85 to 204 minutes to complete the experiment; the average duration was 22 minutes and the pre-determined maximum duration allowed was 45 minutes. Two participants were excluded because the standard deviation of their judgements was lower than 5. All the participants performed thoroughly the typing task, so no participant was excluded according to this criterion. The final sample size was n = 69. Alpha was set to .05 for all analyses.

Judgment. Recency effects were observed for the pleasantness ratings: End stories (M = 39.9, SD = 14.6) were rated as more unpleasant than Middle (M =

[27] As in Experiments 8 and 9, this second memory task was simply used to validate the descriptions typed in by participants; no analyses were run on the selection task only

43.6, *SD* = 14.8) and Start stories (*M* = 45.0, *SD* = 15.5). As in the previous three Experiments, the Control lists were rated as the most pleasant story type (*M* = 64.2, *SD* = 17.0).

A repeated-measures one-way ANOVA revealed that the main effect of Story type (Start, Middle, End, and Control) was significant, $F(3, 204) = 87.6$, $p < .001$, $\eta_p^2 = .56$. Planned comparisons revealed once more that the Control stories were rated as the most pleasant (all *p*s < .001). End stories were rated as less pleasant than Start ($t(68) = 4.5$, $p < .001$, $d = .54$) and Middle stories ($t(68) = 3.6$, $p < .001$, $d = .43$); the difference in pleasantness ratings between Start and Middle stories approached significance, $t(68) = 1.4$, $p = .08$. Figure 4.6 below summarises these findings.

As for the previous two Experiments, no significant difference were noted between the three different versions of the Control lists, $F(2, 66) = 1.4$, $p > .23$.

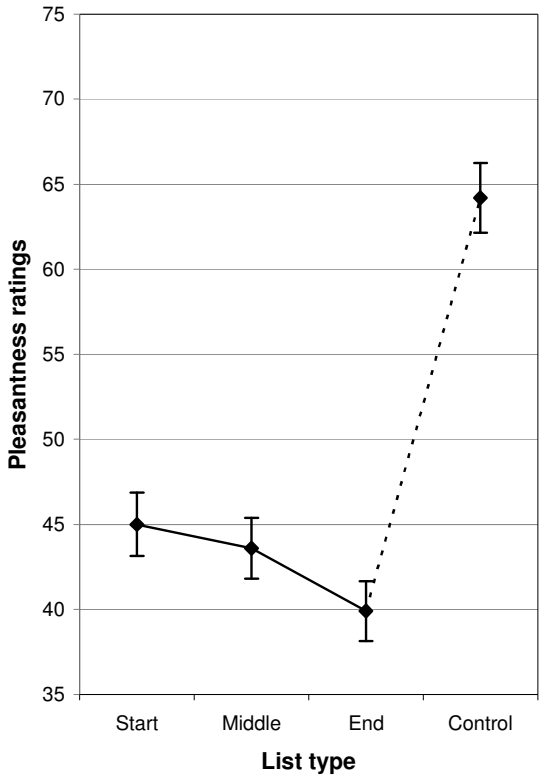


Figure 4.6 Mean pleasantness ratings as a function of story type. Error bars indicate *SEM* (Exp.10)

Memory.[28] Figure 4.7 represents the mean recall proportion as a function of event position and valence. Negative events (M = .69, SD = .31) were better recalled than the neutral ones (M = .32, SD = .23) – this memory advantage being evident across presentation positions. Memory performance for the negative events decreased linearly across presentation positions: Negative events presented in last position (M = .77, SD = .31) were better remembered than those presented in middle positions (M = .69, SD = .30), which in turn were recalled more frequently than the negative events presented at the beginning of the stories (M = .61, SD = .33).

[28] As in the previous two Experiments, the data for Neutral events will refer to the narratives matched for familiarity with the Negative events

A 2 (Valence: negative vs. neutral) × 3 (Position: Start, Middle, and End) repeated measures ANOVA confirmed these observations. The analysis revealed significant main effects of Position ($F(2, 136) = 32.7$, $p < .001$, $\eta_p^2 = .33$) and Valence ($F(1, 68) = 94.6$, $p < .001$, $\eta_p^2 = .58$). The interaction between Valence and Position was significant, too ($F(2, 136) = 9.7$, $p < .001$, $\eta_p^2 = .13$). Follow-up analyses revealed different recall patterns for negative and neutral events: For the former Middle events were better remembered than Start events ($t(68) = 2.1$, $p < .05$, $d = .26$), while this was not true for neutral events, $t(68) = 0.7$, ns.

Most importantly, negative End events were better remembered than Middle ($t(68) = 2.2$, $p < .05$, $d = .28$) and Start negative events ($t(68) = 4.1$, $p < .001$, $d = .49$).

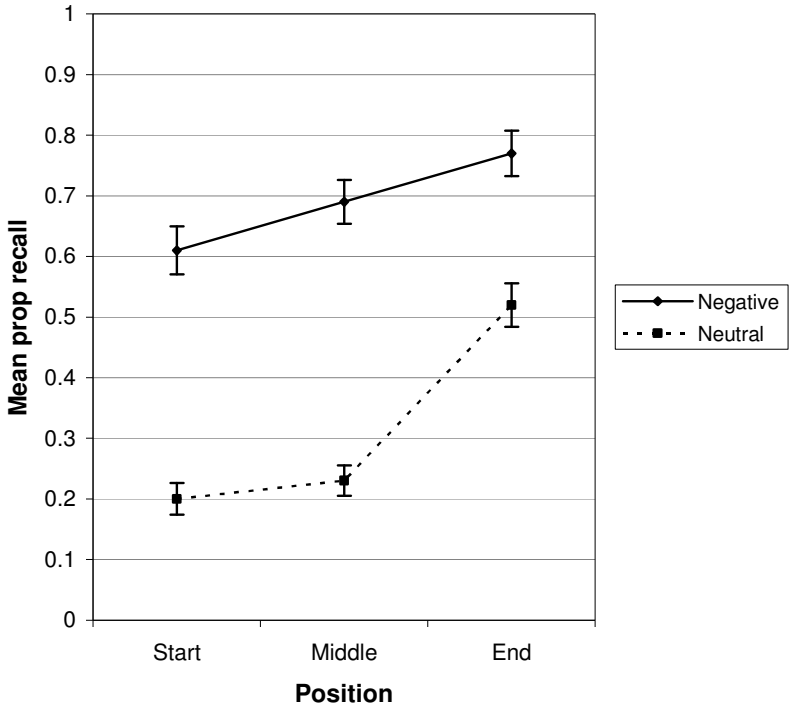


Figure 4.7 Mean recall proportion as a function of event position and valence. Error bars indicate *SEM* (Exp.10)

Memory-Judgment relationships. As for Experiment 9, analyses were run on corrected judgment scores (J'), which represented how much more unpleasant was each story which contained a negative event (Start, Middle, and End) compared to the Control stories. J' scores were again averaged for each participant, according to the negative event presentation position and whether the negative event presented in the story was recalled or not[29,30].

As in Experiment 9, when the negative event was recalled in the memory task, pleasantness ratings for the story *as a whole* were lower (M = -23.5, SD = 17.2) than when it was not recalled (M = -14.7, SD = 14.6). Moreover, this relationship between memory and judgment was relatively constant across story types. Table 4.3 below summarises these findings.

Table 4.3 Mean corrected pleasantness ratings (J') as a function of story type and the negative event being recalled or not (Exp.10)

		Story type		
		Start	Middle	End
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)

[29] This analysis yielded a total of 32.1% missing values, which were missing completely at random as the MRC Little's Test was not significant, χ^2 (73) = 7.6, $p > .55$. Missing values were replaced once again via mean by subject substitution

[30] As the missing values represented a large amount of cases (32.1%) correlational analyses were run as for Experiments 9 in order to confirm the association between memory and judgment measures. The negative event recall proportion correlated significantly with the average pleasantness ratings (r = -.40, $p < .001$ for Start lists; r = -.52, $p < .001$ for Middle lists; r = -.21, $p < .05$ for End lists)

A 2 × 3 within-subjects ANOVA was run with Memory (Negative event recalled Vs. not recalled) and Story type (Start, Middle, and End) as the factors. The significant main effect of Memory confirmed the association between memory and judgment measures, $F(1, 68) = 30.0$, $p < .001$, $\eta_p^2 = .31$. The relationship between memory and judgment was observed regardless of presentation position, as the Story type by Memory interaction only approached significance ($F(1.7, 116.9) = 3.4$, $p = .07$)[31].

Finally the effects of *accessibility* (e.g., Schwarz, 1995; Schwarz & Vaughn, 2002; Tversky & Kahneman, 1973) of negative events on summary assessments were tested. In order to do so, pleasantness ratings were analysed depending on the negative narrative output position (see Table 4.4) – whether it was recalled as first or second response[32].

Table 4.4 Mean corrected pleasantness ratings (J') as a function of negative event recall output position (Exp.10)

		Negative event recall output position		
		Not recalled	Recalled as first response	Recalled as second response
Pleasantness ratings (J')	*M*	-14.7	-24.6	-20.9
	SD	(14.6)	(17.6)	(17.5)

As expected, judgments were more unpleasant when the negative event was recalled as the first response ($M = -24.6$, $SD = 17.6$) as compared to when it was recalled as the second response ($M = -20.9$, $SD = 17.5$). Judgments were

[31] Degrees of freedom were adjusted according to the Greenhouse-Geisser correction as the assumption of Sphericity was not met. Moreover, follow up analyses revealed that pleasantness ratings were lower when the negative event was recalled compared to when it was not recalled, regardless of its presentation position (all *p*s < .001)

[32] This analysis yielded a total of 8.2% missing values, which were once again imputed via mean by subject substitution

clearly most pleasant when participants did not recall the negative event at all (M = -14.7, SD = 14.6).

A one-way ANOVA was run and revealed a significant main effect of negative event output position, $F(2, 136) = 18.7$, $p < .001$, $\eta_p^2 = .22$. Planned comparisons confirmed that negative events recalled as first response were associated with more unpleasant ratings that those recalled as second response $t(68) = 3.1$, $p < .01$, $d = .59$. Obviously, the situations which attracted the most pleasant ratings were those when the negative event was not recalled (both ps < .001).

The results for the present Experiment showed that, when measures are taken to align the information accessed in both memory and judgment tasks, clear associations between recall and judgment measures can be observed. The results for retrospective evaluations produced the same pattern observed in the previous three experiments: the later in the story a negative event was presented, the larger its impact on pleasantness ratings. It seems that participants "look back" at the just-presented story and they are influenced by the mostly presented events.

As in Experiment 9, the memory results mirrored those obtained for judgment. Negative events, which were arguably most distinctive within the stories, were better remembered than neutral items – and this was true across presentation positions. More importantly, memory performance for the negative events increased along presentation positions: the later a negative event happened within the story, the more likely that it was remembered.

The correlational analyses confirmed that memory and judgment measures were associated. First, as predicted by the availability principle (i.e.

availability as *content of recall*; cf. Schwarz & Vaughn, 2002), stories for which the negative event was recalled were rated as more unpleasant than those stories for which it was not recalled. Secondly, *accessibility* (i.e. availability as *ease of recall*; cf. Schwarz & Vaughn, 2002) moderated the impact of the negative event on summary assessments: the most unpleasant ratings were provided when the negative event was recalled as the first response – that is, when it was most easily accessible in memory.

4.6. General Discussion

The general objective of the present series of studies was to test a memory-based approach to retrospective evaluations with stimuli other than lists of words. The results described in the experiments reported in Chapters 2 and 3 highlighted that memory may bias retrospective evaluations through two different dynamics (i.e. heuristics). First, judgments seemed to be based on the *content of recall* (cf. Schwarz & Vaughn, 2002): lists where the negative word was recalled at the memory stage were rated as more unpleasant than those where it was *not* recalled. Second, the influence of a negative item on summary assessment was moderated by its *accessibility* (i.e. its *ease of recall*; cf. Schwarz & Vaughn, 2002; Tversky & Kahneman, 1973); the easier it was to access in memory the negative item, the larger its impact on retrospective evaluations. The same results were obtained with short stories told through slideshows, which arguably represent more cohesive and ecologically valid stimuli than lists of semantically unrelated words. In doing so, we ruled out the possibility that the stimuli non-cohesiveness (i.e. being low in *entitativity*; cf. Campbell, 1958, cited

in McConnell et al., 1997) was the main reason for the memory–judgment relationships observed in Chapters 2 and 3.

However, as for the experiments using word lists, the hypothesis that *on-line judgments* (i.e. the impression formed while experiencing the stimuli; e.g., Hastie & Park, 1986) also influence summary assessments cannot be eliminated. Stories where the negative event was not recalled were still rated as more unpleasant than Control stories, where no negative event was presented at all (by 12.5 units on average in Experiment 9 and 14.7 units in Experiment 10). A potential explanation of this finding invokes both on-line and memory-based processes (see also Chapter 2). When providing summary assessments, participants formed an on-line impression of the story while it was being told; when prompted with the rating task, participants may have adjusted this impression depending on the events that were most easily available in memory.

The present series of studies highlighted two important and novel findings. First, it was shown how summary assessments for slideshow stories produce different patterns compared to lists of words – when the same type of position manipulation was implemented. For word lists, a negative stimulus presented in the first or last positions affected pleasantness ratings to a greater extent than a negative item presented in the middle positions. That is, *both* primacy *and* recency effects were observed for retrospective evaluations. On the other hand, only recency effects were observed for story pleasantness ratings.

A simple memory-based approach to bias in summary assessments can accommodate these findings. The summary assessments were predicted by the memory participants exhibited about the negative and distinctive stimulus inserted within the series. Both primacy and recency effects were observed for

memory for negative words (Experiments 2 and 3) while only recency effects were found for slideshow stories (Experiments 9 and 10). Correlational analyses provided findings which are in accordance with the availability heuristic (Tversky & Kahneman, 1973), according to which retrieval may be the most likely process that bias retrospective judgments.

Another important observation that can be drawn on the basis of the evidence collected in this Chapter is that *the way memory is tested* may have a large repercussion on the memory–judgment relationships. The extent to which recall tasks cue the same information that participants called to mind when providing the judgments may prove to be a key factor in the study of memory and judgment associations (see also Moser, 1992; Shedler & Manis, 1986). In Experiment 8 participants were asked to assess the pleasantness of the stories– which arguably was determined by the pleasantness of the *events* that formed them – but were asked to describe the two *pictures* that come to mind most easily. As the two tasks tapped into different representations of the just-seen stories, different memory and judgment patterns were observed. On the other hand, when during the memory tasks participants were asked to recall the information upon which they most likely formed their judgments, clear associations between memory and judgment measures were observed (Experiments 9 and 10).

In conclusion, it was shown that distinctive moments within a story impact retrospective judgments in a way that is predicted by a memory-based approach to biases in summary assessments. The reported results provide compelling evidence in favour of a viewpoint where remembered episodic

information has a biasing effect on judgment in hindsight even for slideshow stories.

Chapter 5:

General Discussion

5.1. Summary of empirical work

5.1.1. Why are people biased when evaluating the past?

Review of rationale and aims

The aim of the present work was to investigate some of the cognitive processes underlying judgment biases in retrospective evaluations. Retrospective evaluations are simply summary assessments provided in hindsight. Understanding how evaluations of the past are generated is of import as they often constitute the basis of future decisions: How an experience is evaluated affects the decision to repeat (or not repeat) the experience itself or recommend it to others (e.g., Kahneman, 2000a; 2000b). Previous research has shown that when people retrospectively assess life events, they are heavily influenced by specific moments within the episode. For instance, the "Peak-End" rule states that the most intense (the "Peak") and the last moments of an event (the "End") are the best predictors of global assessments, while more holistic features (e.g., event duration and its average quality) are neglected (Fredrickson & Kahneman, 1993). Strikingly, while these judgment biases have been widely documented (e.g., Ariely & Carmon, 2000), a comprehensive cognitive account of *why* these biases arise is somehow missing.

In the present thesis, it was hypothesised that memory may play an important role in biasing retrospective evaluations. It was argued that, as memorability of items increases when they are distinctive (e.g., Hunt, 2006), so the most intense moments of an episode are more readily available in memory. The increased memorability of these moments would in turn affect retrospective

judgments – as they would be overweighed in the judgments of the episode *as a whole*. While covering research areas such as social cognition (e.g., Hogarth & Einhorn, 1992) and behavioural decision making (e.g., Ariely, 1998), the present work addressed the broader debate surrounding the relationships between memory and judgment (e.g., Hastie & Park, 1986). On the one hand, Independence models (e.g., Anderson, 1989) claim that judgments are formed on-line, while the event is being experienced; thus, retrospective evaluations do not depend on the information about the event people hold in memory. On the other hand, several empirical findings have suggested that judgment depends, at least in part, on memory processes (e.g., Schwarz, 1998; Tversky & Kahneman, 1973).

Chapter 2: The role of memory in retrospective evaluations

An experimental paradigm was devised in order to contrast the predictions of different approaches to judgment bias in retrospective evaluations, including the "Peak-End rule" (Fredrickson & Kahneman, 1993), the "Belief-Adjustment" model (Hogarth & Einhorn, 1992), and the "Value-Account" (Betsch, Plessner, Schwieren, & Gütig, 2001). The contribution of memory processes to these biases was investigated in a controlled and systematic way – while using a range of memory measures. The results showed that the impact of a negative item on the pleasantness ratings of word lists was largely influenced by its presentation position. As expected on the basis of primacy *and* recency effects in memory for this type of stimuli (e.g., Murdock, 1962), the most unpleasant lists were the ones where the negative word was presented either at the beginning or at the end. Further analyses supported the association between memory and

judgment measures. For instance, the ease with which a negative word was retrieved from memory (*accessibility*; Schwarz, 1998; Schwarz & Vaughn, 2002; Tversky & Kahneman, 1973) was a predictor of its negative impact on retrospective assessments; the most unpleasant ratings were provided in those situations when the negative item was most easily recalled.

Also, the results from Experiments 2 and 3 provided some insight into the influence of memory task demands on the relationships between memory and retrospective judgment. The results suggested that a challenging memory task, where participants were asked to recall as much information as they could, weakened the relationships between memory and retrospective judgment. On the other hand, the clearest correlations between memory and judgment measures were observed with a novel task, which was introduced to lower the memory task demands.

Chapter 3: The effects of manipulating accessibility

As in Chapter 2 there was some clear evidence of an association between memory and judgment measures, Chapter 3 was attempted to investigate the nature of the relationship further. The experiments run in Chapter 3 offered a stricter test of a memory-based approach to retrospective judgment. Relying on well-established memory phenomena, memory of the to-be-assessed episode was manipulated and judgment predicted *a priori*. For instance, in Experiment 5, negative items which were preceded and followed by a temporal interval were better recalled than non-isolated negative words (temporal isolation effect; Brown, Neath, & Chater, 2007). Accordingly, pleasantness ratings for the lists

including the former type of distinctive items were lower than for the lists where the negative item was not isolated.

Furthermore, a very important methodological issue was addressed. As in most of the studies which investigated memory–judgment relationships (see Hastie & Park, 1986 for a review), a dual task paradigm was used: participants had to both assess and remember the materials they were presented with. It can be argued that when the memory task *precedes* the judgment task, correlations between memory and judgment might be inflated because participants are artificially induced to rely on their memory outputs in order to provide summary assessments (e.g., Schwarz, Strack, & Mai, 1991). For this reason, it was decided to present participants with the opposite task order. However, when the memory task *follows* the judgment task, the observed correlations between memory and judgment might be inflated as well; after all, participants might be accessing the to-be-assessed information from memory at the time they provide the judgment simply because they know they have to perform a memory task subsequently. Experiment 6 ruled out this possibility, as availability and accessibility biases were observed even when participants were presented with a *surprise* memory task.

Chapter 4: Retrospective evaluations of stories

Chapter 4 asked whether the above findings could be replicated with different stimuli. Participants attended to and evaluated short stories told through slideshows, which arguably represent more ecologically valid and *cohesive* episodes than lists of semantically unrelated words. Moreover, it was more likely that participants could develop a representation of the episode based on "gist"

(e.g., Brainerd & Reyna, 1992). Biases in retrospective evaluations were different from those observed in Chapters 2 and 3 – but were mostly accounted for by a memory-based approach. The most unpleasant stories were those that ended with a negative event, which was more readily available in memory than negative events happening at any other point during the story.

Chapter 4 also provided some insight into the nature of the representation accessed in memory in order to provide summary assessments. It was suggested that the alignment between the information cued by the judgment and memory task is important. When the memory and the judgment tasks cued different features of the episode, no associations between memory and judgment were observed (Experiment 8). When the recall task tapped more into the content that was likely to be called to mind – at least according to a memory-based approach – in order to provide the evaluations, memory–judgment correlations were again observed (Experiments 9 and 10).

Conclusions

The present thesis focused on the biases in retrospective evaluations and sought to empirically address the question of *why* these biases arise. The results suggest that memory processes may be at the base of the overweighing of specific moments when retrospectively assessing an experience.

5.2. A memory-based approach

Considerations about the retrospective judgment patterns

While *both* primacy *and* recency effects were observed for pleasantness ratings of lists of words, recency effects *only* were observed for assessments of slideshow stories. Despite the fact that the position of a negative and distinctive stimulus was manipulated in the same manner for lists of words (Chapter 2) and slideshow stories (Chapter 4), its effects on summary assessments differed. Lists of words which included a negative item either in first or last positions were more unpleasant than those lists where the negative item was presented in the middle positions (i.e., primacy and recency effects). On the other hand, when a negative event was inserted within a short story told through slideshows, it affected retrospective judgment the most when it was presented at the end (recency effects).

This observation has implications for previous accounts of retrospective evaluations. According to the "Peak-End" rule (Fredrickson & Kahneman, 1993), retrospective evaluations are best predicted by an average of the quality experienced during the most intense moment of the episode (the "Peak") and its final moments (the "End"). As such, this general formulation of the "Peak-End" rule could explain the results observed for slideshow stories. When a negative event (the "Peak") was presented at the end of a story, its effects on judgment were the largest – "Peak" and "End" coincided. For lists of words, the consistently observed primacy effects in retrospective judgments cannot be predicted by the "Peak-End" rule. The negative word inserted in the to-be-assessed lists is the most likely candidate for the "Peak" or more distinctive item. From the perspective of the "Peak-End" rule, there is no reason to expect an increased weight in the summary assessment when the negative item is presented

in first position. It follows that, based on the Peak-End rule, the pleasantness ratings for those lists where the "Peak" was presented in the first position should be similar to the ratings observed when the negative item was in the middle positions. It is acknowledged here that typical "Peak-End" effects have been observed mostly with experiential and continuous stimuli like mechanical pressure (Ariely, 1998) and movie clips (Fredrickson & Kahneman, 1973) – contrary to the discrete nature of the stimuli used in the present work[33]. Furthermore, recent evidence has shown that *segmentation* can lessen the biases predicted by the "Peak-End" rule (e.g., Ariely & Zauberman, 2000; 2003). Nonetheless, it is important to notice that both lists of words and slideshows were presented to participants in a discrete manner, as both items *and* slides were separated by a temporal gap. Arguably then, what differed between the two types of stimuli was not their discrete nature *per se*, but the cohesiveness of the *representation* called in mind in order to evaluate the events.

The present series of findings cannot be accommodated by the "Value-Account" (Betsch et al., 2001). According to this summative approach to retrospective judgments, what matters the most is the *amount* of negative and positive information within a to-be-assessed sequence – no specific hypotheses are put forward relative to the *position* of such information.

In the "Belief-Adjustment" model (Hogarth & Einhorn, 1992), the type of bias observed in retrospective judgments depends on how the to-be-assessed information is mentally processed. In Step-by-Step processing, the judgment of the series is adjusted after each piece of information; thus, recency effects are predicted, as the last encountered information is weighed as much as all the

[33] However, recent evidence has highlighted "Peak-End" effects also for stimuli which were discrete, semantic and relative short in nature (e.g., Do et al., 2008; Langer et al., 2005)

aggregated information that preceded it. Instead, End-of-Sequence processing involves one single adjustment, where the initial impression is integrated with all the succeeding information; this leads to primacy effects in retrospective judgment. The type of processing used depends both on stimuli-related features – like their length and complexity – and the familiarity with the task. As neither changed substantially between trials, it could be argued that the "Belief-Adjustment" model would have predicted *either* primacy *or* recency effects for the current paradigm.

The association between memory and judgment measures

The evidence presented in this thesis supports theories and approaches which include memory processes amongst the cognitive precursors of judgment biases (e.g., Dougherty, Gettys, & Ogden, 1999; Schwarz, Bless, Wanke, & Winkielman, 2003; Stewart, Chater, & Brown, 2006; Tversky & Kahneman, 1973; Weber & Johnson, 2006). Despite some experimental dissociations between memory and judgment (e.g., Experiment 4), most of the studies reported seem in favour of an association between them.

Generally, pleasantness ratings were lower in those situations in which the negative and distinctive stimulus (either a word or an event) was recalled in the memory component of the task – as compared to when it was *not* recalled. This finding seem to support the influence of *availability*, and the claim that judgments were based on *what* comes to mind (e.g., Higgins, 1996; Reyes et al., 1980).

Moreover, the most unpleasant ratings were observed when the negative word (or event) was most easily recalled – if output order is considered an

acceptable indicator of ease of recall. Thus, it also seemed that *accessibility*, operationalised as ease of retrieval, moderated retrospective judgments.

It could be argued that during all the experiments described in the present thesis participants could form on-line judgment (e.g., Anderson, 1996; Anderson & Hubert, 1963; Hastie & Park, 1986); that is, they could evaluate the stimuli as they were presented to them. According to the "Independence models" (cf. Hastie & Park, 1986), this situation should have led to dissociations between memory and judgment measures most of the time. After all, from this perspective, participants simply had to rely on the impression of the stimuli they formed while attending to them – retrieving information about the episodes was not necessary in order to evaluate them in hindsight. While the role of on-line judgment cannot be completely ruled out – and it will be addressed below – the observed associations between memory and judgment seem to go against the predictions of Independence models.

The present thesis also showed that it is possible to *predict* retrospective evaluations on the basis of memory. The memorability of distinctive moments within an episode was manipulated, through serial position (Experiments 1, 2, and 3) and isolation effects (Experiment 5). As a consequence, the impact of those moments on summary assessments was affected: Their effect was greatest when they were most easily available in memory. These results are of a certain importance as, to our knowledge, few studies in the literature tested *a priori* hypotheses on retrospective evaluations depending on the memory manipulation that was implemented (e.g., Gabrielcick & Fazio, 1984; Lewandowsky & Smith, 1983). Moreover, the few that did so mainly involved *frequency* judgments, whose underlying cognitive processes, as discussed in Chapter 1, may differ

substantially from those underpinning *quality* judgments (i.e., *estimation* tasks; cf. Hogarth & Einhorn, 1992; p. 9).

Overall, the retrospective evaluations participants provided closely mirrored their memory of the to-be-assessed stimuli. Primacy *and* recency effects were observed for the recall of negative words – and primacy and recency effects were observed for the pleasantness ratings of lists of words. In Chapter 4, recency effects *only* were observed for both the recall of negative events and the pleasantness ratings for those stories which encompassed them. Thus, the observed biases in retrospective evaluations can be accounted for by the biases observed for the recall of the negative stimulus within the to-be-assessed episode.

Why did the memory biases differ depending on the stimuli used? One explanation may lie in the effects of *rehearsal*. Arguably, participants may have engaged more readily in some rehearsal when presented with word lists – as compared to when they attended to short slideshows. In turn, the *primacy* effects observed in Chapter 2 may have been due to the larger number of rehearsals for the early items presented, as compared to later items. Empirical evidence has indeed suggested that the probability of recalling early items can be influenced by the number of rehearsals they receive, which are usually more numerous than for later items (Rundus, 1971; cited in Tan & Ward, 2000). On the other hand, slideshow stories included more complex stimuli, where each event was described in a multi-modal manner, through a picture and a description which was both typed and spoken. Moreover, slideshows represented more cohesive experiences than lists of words, as for the former it was undoubtedly easier for participants to form a gist (Brainerd & Reyna, 1992). Combined, these features

may induced participants to rehearse the events in the stories to a lesser extent – or not at all; as a consequence, when "looking back" at the just-attended story, its events were more readily available in memory the later they happened – which resulted in the observed recency effects and very little primacy.

Factors that moderate the memory–judgment relationships

The data of the present thesis have provided some further interesting findings about the relationships between memory and judgment. In Chapter 2, the clearest associations between memory and judgment measures were observed when the memory task was the least exacting. A memory task was devised and involved participants having to recall the two items that came to mind most readily, and any other item they may remembered subsequently. The major difference with a *free recall* task is that in the latter participants are asked to recall *as many items as they can remember* – which undoubtedly places greater demands on participants. Thus, it appears that the cognitive demands associated with the memory task can affect the correlations between memory and judgment. These findings are in the line with previous research which 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 (Schwarz & Vaughn, 2002; see also Higgins, 1996). Similarly, Kitayama and Burnstein (1989) have argued that typical recall tasks, which require participants to perform an exhaustive search in memory, can hinder the underlying associations between memory and judgment. As people seem to base their evaluations on partial information – on the elements they can retrieve and that are most easily

accessible – it is perhaps not surprising that previous research which used standard and demanding memory tasks struggled to find correlations between memory and judgment measures (e.g., Anderson & Hubert, 1963; Hastie & Park, 1986; Shedler & Manis, 1986).

In line with the above, it is suggested that the *way* memory is tested can largely determine whether memory and judgment measures will be associated (see also Moser, 1992). More specifically, the underlying relationship between memory and judgment is most likely to be detected when the memory task *cues* the same information as that which the judgment was most likely to be based upon. On the other hand, when retrospective evaluations are obtained in one context and memory for the materials is assessed in a different manner, the above correlations may be hindered. For instance, in Experiment 8, where participants were asked to rate the pleasantness of the *stories* but were asked to remember the *pictures* that most readily came to mind, memory and judgment measures were not associated. On the other hand, when participants were asked to rate the stories and then to retrieve the two *moments* that most readily came to mind, correlations between memory and judgment were observed again (Experiment 10).

It can be concluded that, in order to thoroughly assess the role memory plays in biasing retrospective evaluations, care has to be taken to cue the appropriate information at the time memory is probed; the memory test has to adequately reflect the what was relied upon at the time of judgment.

Could memory–judgment associations be judgment-driven?

A few theoretical approaches hold that the correlations between memory and judgment measures are a result of the influence of evaluative processes on memory. Here are a few illustrations of this idea. The "Incongruent-based encoding" hypothesis (Hastie, 1984; see also the "Biased encoding" account; Alba & Hasher, 1983) proposes that judgment processes bias memory during the encoding phase. Information that is *incongruent* with an initial judgment (i.e., a sort of "first impression" of the to-be-evaluated stimuli) is better remembered than congruent information. In a study by Hastie and Kumar (1979), participants were presented with behaviour descriptions about a hypothetical character; later, they were asked to judge her personality. During a following memory test phase, the behaviours that were incongruent with the personality appraisal were better remembered. The authors argued that incongruent behaviours received special processing during the encoding phase, which led to their observed memory advantage. The "Biased retrieval" hypothesis (Leamer, 1974), despite predicting the bias to happen at a different memory stage, still holds that it is judgment which causes memory, and not the other way around.

One of the main findings of Chapter 3 seems to rule out the above described judgment-driven view, at least for the present experimental paradigm and stimuli. In Experiment 5, lists which included *isolated* negative items were rated as more unpleasant than those where the negative words were not isolated. It was argued that, by isolating the negative items, their memorability increased; as a consequence, the effects of negative items on the pleasantness ratings were larger. The opposite argument – which would explain the increased memorability of isolated items on the basis of the decreased pleasantness of the lists which

included them – seems to struggle more to account for the following finding. The decrease in pleasantness ratings for lists which included an isolated negative item was observed *only* when the negative words were recalled. When a negative item was not recalled, its isolation did not moderate retrospective judgments, as these were relatively low regardless of isolation.

Relevance to models of judgment and decision making

The present findings are also relevant to recent models which relate memory and judgment processes in order to explain decision-making phenomena (e.g., Dougherty et al., 1999; Stewart et al., 2006). In their Decision-by-Sampling (hereafter DbS) model, Stewart and colleagues (2006) propose that, when facing a choice, people retrieve from memory instances similar to those at hand (i.e., the *decision sample*). To make a decision, people rank the present option within the decision sample; the outcome of this comparison will determine the subjective value of the to-be-evaluated stimulus. A direct implication of DbS is that the *size* of the retrieved decision sample can affect judgments and choices. Working memory, which in this case determines how large the decision sample retrieved is, plays a central role in moderating the extent to which specific decision-making biases are observed. For instance, *temporal discounting*, i.e. preferring smaller gains because they are nearer in time, is greater when higher demands are placed on working memory (Hinson, Jameson, & Whitney, 2003). Also, Dougherty and Hunter (2003a; 2003b) showed that people with higher memory spans displayed less *subadditivity bias* in probability judgments; that is, they were less likely to judge the probability of a whole to be *smaller* than the probability of its components. However, recent evidence has suggested that

individual differences in working memory capacity might not be sufficient to explain the wide array of decision-making biases (e.g., *framing effects* and *overconfidence bias*; Bruza, Welsh, & Navarro, 2008).

In light of the findings of the present thesis, it could be suggested that incorporating the notion of *accessibility* into DbS may improve its predictive power. Simply put, when making a decision, it is not only *what* people retrieve from memory (and its associated probability of being retrieved; Stewart et al., 2006) that matters; it is also important *how easily* each bit of information in the decision sample is recollected. Indeed, even people with greater working memory spans who retrieve larger decision samples, are biased in their decisions to some extent, possibly because they *overweigh* the information which they retrieved most easily. Previous research has shown that people tend to base their judgments on the subset of information that comes to mind most readily (e.g., Bodenhausen & Wyer, 1987; Higgins, 1996). Some results of the present thesis seem to indicate that, even within this subset, people do not equally weigh the information; rather, they seem less influenced by this information as its ease of retrieval decreases.

On-line anchoring and memory-driven adjustment

A consistent result of Chapters 2 and 3 was that pleasantness ratings were lower for those lists where the negative item was presented *but not recalled* – as compared to those lists where no negative words were presented at all (the "Control" lists). This was also so for the slideshow stories in Chapter 4.

These results do not allow us to discard the role of one of the theoretical pillars of "Independence" models, i.e. *on-line judgment* (e.g., Anderson, 1989;

Betsch et al., 2001). Simply put, it can be argued that participants adjusted their pleasantness ratings after each word (or slide) was encountered. As soon as a negative word was processed, it decreased the (on-line) pleasantness rating; consequently, it did not matter whether the negative item was available at the time the retrospective evaluation task was prompted – and for that matter, during the memory task, either. Indeed, in the present thesis, none of the results can preclude the possibility that participants did not form an impression of the episode as they were attending to it.

Despite the above compelling argument, two alternative explanations are considered here. First, it could be that at the time of retrospective judgment, the negative item was available. However, at the point of recall, in the memory task, that item was no longer accessible. This would imply that the negative item had an impact on the judgment, but was nevertheless not retrieved during the memory component of the task. Situational factors may have led to this dissociation; for instance, output interference could have hindered the likelihood of recalling the negative item within the memory task more on some trials than others (Nairne, 1990).

A second explanation would combine on-line and memory-driven processing. Retrospective evaluations can be the result of an on-line impression and the adjustment that follows when a summary assessment is produced. This adjustment would be driven by the information about the just-experienced event that comes to mind most readily. The latter phase would serve as a form of "verification" of the on-line impression. This "hybrid" approach would go some way towards explaining the findings of the present thesis. First, it could explain why a negative stimulus affected retrospective evaluations even when it was not

recalled. That is, people evaluate episodes as they experience them. Second, it could explain the set of results that consistently indicated associations between memory and judgment.

5.3. Future research

Suggestions for future research

Because of the dual nature of the task and the time lag between the judgment and memory tasks, the representation of the stimuli accessed at the time of the two tasks may have differed. This inconsistency – despite the attempts to minimise it – may have slightly altered the overall picture of the relationships between memory and judgment. The following scenario (previously described) represents an example. A negative item is retrieved from memory at the time of judgment and it causes a decrease in the pleasantness rating of the list which included it. However, because of one of many possible factors (e.g., *output interference*; Nairne, 1990) the same item is not retrieved at the later recall stage. This may have resulted in lists, whose negative item was not recalled, to be rated as more unpleasant than lists which contained no negative item. In order to test this hypothesis, a follow-up memory task could be administered after a conventional recall task. This memory task would call upon implicit recognition measures and it would allow us to ascertain with a higher degree of certainty whether a non-recalled negative item was available at the time of judgment.

Further studies could be run in order to validate the conclusion that judgment biases can be predicted *a priori* on the basis of memory manipulations. The double manipulation used in Experiment 5, involving a distinctive colour and temporal isolation, was, after all, rather manifest. A further and more subtle experimental manipulation would be to control the delay between stimulus presentations and evaluations – and to predict judgments according to well-known memory phenomena associated to differential delays (see Neath & Surprenant, 2003). As an alternative, a more ecologically valid and less overt memory manipulation is *priming*, which could be applied to slideshow stories. For example, before viewing the slideshow stories, participants could be presented with a word-stem completion task, during which some target words would be primed. Then, participants would be presented with two types of slideshow stories, equated for unpleasantness on the basis of normative databases (e.g., Bradley & Lang, 1999; Lang, Bradley, & Cuthbert, 1999). The first type of slideshow would include events where a central word has been previously primed; the other slideshows would still include a negative event, but none of its items were previously primed. It is predicted that the former type of slideshows, despite being normatively equally unpleasant, will be rated as the most unpleasant because of the increased *accessibility* of its negative event.

5.4. Conclusion

Biases in retrospective evaluations are observed when the to-be-evaluated information is presented sequentially or the to-be-assessed event unfolds over time. Thus, their occurrence ranges across several domains, including consumer

behaviour (Russo, Meloy, & Medvec, 1998), economics (Langer et al., 2005) and medical procedures (e.g., Redelmeier & Kahneman, 1996). Previous research has established that specific moments within an experience are overweighed in the summary assessment people provide in hindsight (e.g., Fredrickson & Kahneman, 1993; Hogarth & Einhorn, 1992).

The present thesis has drawn on the memory literature in order to approach the question of *why* specific moments determine summary assessments. The findings suggest that, when assessing an event in retrospect, the memory of the event is accessed and what is retrieved impacts the judgment in hindsight. In addition, it was also shown that the *ease* with which information is brought to mind moderates its impact on retrospective evaluations of affective episodes – in line with what has been reported for frequency judgments (e.g., Schwarz, 1998; Tversky & Kahneman, 1973). Finally, the results suggest that judgments in hindsight can be manipulated. By hindering the memorability of negative information it was possible to reduce its impact on retrospective assessments. These results are at odds with the predictions of "Independence" theories (e.g., the "Value-Account", Betsch et al., 2001) which argue that retrospective evaluations are solely based on on-line judgments. Although it is likely that people form on-line evaluations as events unfold – especially if they are forewarned of the upcoming assessment task – the results reported in this thesis strongly suggest that the most memorable moments of the episode are brought to mind at the time of the judgement and either determine the judgment or significantly bias it.

Bibliography

Alba, J. W., & Hasher, L. (1983). Is memory schematic? *Psychological Bulletin, 93,* 203–231.

Aldrovandi, S., Poirier, M., & Ayton, P. (2008). *The role of memory in retrospective evaluations*. Manuscript submitted for publication.

Anderson, N. H. (1981). *Foundations of information integration theory.* New York: Academic Press.

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. (1996). *A functional theory of cognition*. Mahwah, NJ: Lawrence Erlbaum Associates.

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. (1998). Combining experiences over time: The effects of duration, intensity changes & on-line measurements on retrospective pain evaluations. *Journal of Behavioural Decision Making*, *11,* 19–45.

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science, 12,* 157–162.

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

Ariely, D., Kahneman, D., & Loewenstein, G. (2000). Joint comment on "When does duration matter in judgment and decision making?" (Ariely & Loewenstein, 2000). *Journal of Experimental Psychology: General, 129*, 524–529.

Ariely, D., & Loewenstein, G. (2000). When does duration matter in judgement and decision making? *Journal of Experimental Psychology: General, 129,* 508–523.

Ariely, D., & Zakay, D. (2001). A timely account of the role of duration in decision making. *Acta Psychologica, 108,* 187–207.

Ariely, D., & Zauberman, G. (2000). On the making of an experience: The effects of breaking and combining experiences on their overall evaluation. *Journal of Behavioral Decision Making, 13*, 219–232.

Ariely, D., & Zauberman, G. (2003). Differential partitioning of extended experiences. *Organizational Behavior and Human Decision Processes, 91,* 128–139.

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

Ashgari, A., & Nicholas, M. K. (2004). Pain during mammography: The role of coping strategies. *Pain, 108,* 170–179.

Baines, C. J., To, .T., & Wall, C. (1990). Women's attitudes to screening after participation in the National Breast Cancer Screening Study: A questionnaire survey. *Cancer, 65,* 1663–1669.

Banaji, M. R., Hardin, C., & Rothman, A. J. (1993). Implicit stereotyping in person judgment. *Journal of Personality and Social Psychology, 65,* 272–281.

Bargh, J. A., & Thein, R. D. (1985). Individual construct accessibility, person memory, and the recall-judgment link: The case of information overload. *Journal of Personality and Social Psychology, 49,* 1129–1146.

Barrett, L. F. (1997). The relationships among momentary emotion experiences, personality descriptions, and retrospective ratings of emotion. *Personality & Social Psychology Bulletin, 23,* 1100–1110.

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology.* Cambridge, England: Cambridge University Press.

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.

Beauvois, J., & Cambon, L. (1998). On-line vs. Memory-based orientations and self-report validity. *European Journal of Social Psychology, 27,* 611–616.

Bedell, S. E., Delbanco, T. L., Cook, E. F., & Epstein, F. H. (1983). Survival after cardiopulmonary resuscitation in the hospital. *The New England Journal of Medicine, 309,* 569–576.

Benjamin, A. S., & Bjork, R. A. (1996). Retrieval fluency as a metacognition index. In L. M. Reder (Ed.), *Implicit memory and metacognition* (pp. 309–338). Mahwah, NJ: Lawrence Erlbaum Associates.

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.

Bizer, G. Y., Tormala, Z. L., Rucker, D. D., & Petty, R. E. (2006). Memory-based versus on-line processing: Implications for attitude strength. *Journal of Experimental Social Psychology, 42,* 646–653.

Bodenhausen, G. V., & Wyer, R. S. (1987). Social cognition and social reality: Information acquisition and use in the laboratory and the real world. In H. J. Hippler, N. Schwarz, & S. Sudman (Eds.), *Social information processing and survey methodology* (pp. 6–41). New York: Springer-Verlag.

Bond, S. D., Carlson, K. A., Meloy, M. G., Russo, J. E., & Tanner, R. J. (2007). Information distortion in the evaluation of a single option. *Organizational Behavior and Human Decision Processes, 102,* 240–254.

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.

Brainerd, C. J., & Reyna, V. F. (1990). Gist is the grist: Fuzzy-trace theory and the new intuitionism. *Developmental Review, 10,* 3–47.

Brainerd, C. J., & Reyna, V. F. (1992). Explaining "memory-free" reasoning. *Psychological Science, 3,* 332–339.

Brainerd, C. J., & Reyna, V. F. (1993). Memory independence and memory interference in cognitive development. *Psychological Review, 100,* 42–67.

Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review, 114*, 539–576.

Bruine de Bruin, W. (2005). Save the last dance for me: Unwanted serial position effects in jury evaluations. *Acta Psychologica, 118,* 245–260.

Bruine de Bruin, W., & Keren, G. (2003). Order effects in sequentially judged options due to the direction of comparison. *Organizational Behavior and Human Decision Processes, 92*, 91–101.

Bruza, B., Welsh, M. & Navarro, D. (2008). Does memory mediate susceptibility to cognitive biases? Implications of the Decision-by-Sampling theory. In V. Sloutsky, B. Love, & K. McRae (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society.*

Campbell, D. T. (1958). Common fate, similarity, and other indices of the status of aggregates of persons as social entities. *Behavioral Science, 3,* 14–25.

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.

Clark, A. E., & Georgellis, Y. (2004). Kahneman meets the quitters: Peak-end behaviour in the labour market. Unpublished manuscript, DELTA.

Claypool, H. M., & Carlston, D. E. (2002). The effects of verbal and visual interference on impressions: An associated-systems approach. *Journal of Experimental Social Psychology, 38,* 425–433.

Cohen, J. (1988). *Statistical power analysis for behavioral science (2nd edition).* Hillsdale, NJ: Lawrence Erlbaum Associates.

Cojuharenco, I., & Ryvkin, D. (2008). Peak-End rule versus average utility: How utility aggregation affects evaluations of experiences. *Journal of Mathematical Psychology, 52,* 328–337.

Collins, D. J., & Shanks, D. R. (2002). Momentary and integrative response strategies in causal judgment. *Memory & Cognition, 30,* 1138–1147.

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

Combs, B., & Slovic, P. (1979). Newspaper coverage of causes of death. *Journalism Quarterly, 56,* 837–849.

Conrad, R., & Hull, A. J. (1968). Input modality and the serial position curve in short-term memory. *Psychonomic Science, 10,* 135–136.

Corballis, M. C. (1966). Rehearsal and decay in immediate recall of visually and aurally presented items. *Canadian Journal of Psychology, 20,* 43–51.

Crowder, R. G. (1970). The role of one's own voice in immediate memory. *Cognitive Psychology, 1,* 157–178.

Crowder, R. G. (1986). Auditory and temporal factors in the modality effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 268–278.

Custer, R. & Aarts, H. (2003). On the role of processing goals in evaluative judgments of environments: Effects on judgement-memory relations. *Journal of Environmental Psychology, 23,* 289–299.

Diehl, K., & Zauberman, G. (2005). Searching ordered sets: Evaluations from sequences under search. *Journal of Consumer Research, 31,* 824–832.

Diener, E., Wirtz, D., & Oishi, S. (2001). End effects of rated life quality: The James Dean effect. *Psychological Science, 12,* 124–128.

Do, A. M., Rupert, A. V., & Wolford, G. (2008). Evaluations of pleasurable experiences: The Peak-End rule. *Psychonomic Bulletin & Review, 15,* 96–98.

Dougherty, M. R. P. (2001). Integration of the ecological and error models of overconfidence using a multiple-trace memory model. *Journal of Experimental Psychology: General, 130,* 579–599.

Dougherty, M. R. P., & Franco-Watkins, A. M. (2003). Reducing bias in frequency judgment by improving source monitoring. *Acta Psychologica, 113,* 23–44.

Dougherty, M. R. P., Gettys, C. F., & Ogden, E. E. (1999). MINERVA-DM: A memory processes model for judgments of likelihood. *Psychological Review, 106,* 180–209.

Dougherty, M. R. P., Gronlund, S. D., & Gettys, C. F. (2003). Memory as a fundamental heuristic for decision making. In S. L. Schneider, & J. Shanteau (Eds.), *Emerging perspectives on judgment and decision research* (pp. 125–164). New York: Cambridge University Press.

Dougherty, M. R. P., & Hunter, J. (2003a). Hypothesis generation, probability judgment, and individual differences in working memory capacity. *Acta Psychologica, 113,* 263–282.

Dougherty, M. R. P., & Hunter, J. (2003b). Probability judgment and subadditivity: The role of working memory capacity and constraining retrieval. *Memory & Cognition, 31,* 968–982.

Dougherty, M. R. P., & Sprenger, A. (2006). The influence of improper sets of information on judgment: How irrelevant information can bias judged probability. *Journal of Experimental Psychology: General, 135,* 262–281.

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.

Elwood, M., McNoe, B., Smith, T., Bandaranavake, M., & Doyle, T. C. (1998). Once is not enough: Why some women do not continue to participate in a breast cancer screening programme. *The New Zealand Medical Journal, 111,* 180–183.

Fiske, S. T., Taylor, S. E., Etcoff, N. L., & Laufer, J. K. (1979). Imaging, empathy, and causal attribution. *Journal of Experimental Social Psychology, 15,* 356–377.

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.

Gabrielcick, A., & Fazio, R. H. (1984). Priming and frequency estimation: A strict test of the availability heuristic. *Personality and Social Psychology Bulletin, 10,* 85–89.

Glenberg, A. M. (1984). A retrieval account of the long-term modality effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 16–31.

Hands, D. S., & Avons, S. E. (2001). Recency and duration neglect in subjective assessment of television picture quality. *Applied Cognitive Psychology, 15,* 639–657.

Hanita, M., Gavanski, I., & Fazio, R. H. (1997). Influencing probability judgments by manipulating the accessibility of sample spaces. *Personality and Social Psychology Bulletin, 23,* 801–813.

Hastie, R. (1981). Schematic principles in human memory. In E. T. Higgins, C. P. Herman, & M. P. Zanna (Eds.), *Social cognition: The Ontario Symposium (Vol. 1)* (pp. 39–88). Hillsdale, NJ: Erlbaum.

Hastie, R. (1984). Causes and effects of causal attribution. *Journal of Personality and Social Cognition, 46,* 44–56.

Hastie, R., & Kumar, P. A. (1979). Person memory: Personality traits as organizing principles in memory for behaviours. *Journal of Personality and Social Psychology, 37,* 25–38.

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

Hastie, R., & Pennington, N. (1989). Notes on the distinction between memory-based versus on-line judgments. In J. N. Bassili (Ed.), *On-line cognition in person perception* (pp. 1–19). Hillsdale: Erlbaum.

Heuer, F., & Reisberg, D. (1990). Vivid memories of emotional events: The accuracy of remembered minutiae. *Memory & Cognition, 18,* 496–506.

Higgins, E. T. (1989). Knowledge accessibility and activation: Subjectivity and suffering from unconscious sources. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 75–123). New York: Guildford.

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.

Higgins, E. T., Rholes, W. S., & Jones, C. R. (1977). Category accessibility and impression formation. *Journal of Experimental Social Psychology, 13,* 141–154.

Hinson, J. M., Jameson, T. L., & Whitney, P. (2003). Impulsive decision making and working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 298–306.

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

Hsee, C. K., Abelson, R. P., & Salovey, P. (1991). The relative weighing of position and velocity in satisfaction. *Psychological Science, 2,* 263–266.

Hulme, C., Neath, I., Stuard, G., Shostak, L., Surprenant, A. M., & Brown, G. D. A. (2006). The distinctiveness of the word-length effect. *Journal of Experimental Psychology: Learning, Memory and Cognition, 32,* 586–594.

Hunt, R. R. (1995). The subtlety of distinctiveness: What Von Restorff really did. *Psychonomic Bulletin and Review, 2,* 105–112.

Hunt, R. R. (2006). The concept of distinctiveness in memory research. In R. R. Hunt & J. B. Worthen (Eds.), *Distinctiveness and Memory* (pp. 3–25). New York: Oxford University Press.

Hunt, R. R., & Lamb, C. A. (2001). What causes the isolation effect? *Journal of Experimental Psychology: Learning, Memory and Cognition, 27,* 1359–1366.

Jochen, R., & Cropanzano, R. (2007). Evaluating dynamic performance: The influence of salient Gestalt characteristics on performance ratings. *Journal of Applied Psychology, 92,* 490–499.

Kahneman, D. (2000a). Experienced utility and objective happiness: A moment-based approach. In D. Kahneman, & A. Tversky (Eds.), *Choices, Values and Frames* (pp. 673–692). New York: Cambridge University Press.

Kahneman, D. (2000b). Evaluations by moments: Past and future. In D. Kahneman, & A. Tversky (Eds.), *Choices, Values and Frames* (pp. 693–708). New York: Cambridge University Press.

Kahneman, D., Fredrickson, B. L., Schreiber, C. A., & Redelmeier, D. A. (1993). When more pain is preferred to less: Adding a better end. *Psychological Science, 4,* 401–405.

Kahneman, D., & Thaler, R. H. (2006). Utility maximization and experienced utility. *Journal of Economic Perspectives, 20,* 221–234.

Kahneman, D., Wakker, P., & Sarin, R. (1997). Back to Bentham? Explorations of experienced utility. *The Quarterly Journal of Economics, 112,* 375–405.

Kardes, F. R. (1986). Effects of initial product-judgments on subsequent memory-based judgments. *Journal of Consumer Research, 13,* 1–11.

Kardes, F. R., & Kalyanaram, G. (1992). Order-of-entry effects on consumer memory and judgment: An information integration perspective. *Journal of Marketing Research, 29,* 343–357.

Kashima, Y., & Kerekes, A. R. Z. (1994). A distributed memory model of averaging phenomena in person impression formation. *Journal of Experimental Social Psychology, 30,* 407–455.

Kemp, S., Christopher, D. B. B., & Furneaux, L. (2008). A test of the Peak-End rule with extended autobiographical events. *Memory & Cognition, 36,* 132–138.

Kintsch, W., Welsch, D., Schamlhofer, F., & Zimmy, S. (1990). Sentence memory: A theoretical analysis. *Journal of Memory and Language, 29,* 133–159.

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.

Knoedler, A. J., Hellwig, K. A., & Neath, I. (1999). The shift from recency to primacy with increasing delay. *Journal of Experimental Psychology: Learning, Memory and Cognition, 25*, 474–487.

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review, 100,* 609–639.

Koriat, A. (1995). Dissociating knowing and the feeling of knowing: Further evidence for the accessibility model. *Journal of Experimental Psychology: General, 124,* 311–333.

Koriat, A., & Levy-Sadot, R. (2001). The combined contributions of the cue-familiarity and accessibility heuristics to feeling of knowing. *Journal of Experimental Psychology: Learning, Memory and Cognition, 27,* 34–53.

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.

Leamer, E. E. (1974). False models and post-data model constructions. *Journal of the American Statistical Association, 69,* 122–131.

Lewandowsky, S., Duncan, M., & Brown, G. D. A. (2004). Time does not cause forgetting in short-term serial recall. *Psychonomic Bulletin & Review, 11,* 771–790.

Lewandowsky, S., & Smith, P. W. (1983). The effect of increasing the memorability of category instances on estimates of category size. *Memory & Cognition, 11,* 347–350.

Libkuman, T. M., Otani, H., Kern, R. P., Viger, S. G., & Novak, N. (2007). Multidimensional normative ratings for the International Affective Picture System. *Behavior Research Methods, 39,* 326–334.

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.

Lichtenstein, S., Slovic, P., Fischhoff, B., Layman, M., & Combs, B. (1979). Judged frequency of lethal events. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 551–578.

Liersch, M. J, & McKenzie, C. R. M. (2009). Duration neglect by numbers – And its elimination by graphs. *Organizational Behavior and Human Decision Processes, 108,* 303–314.

Lodge, M., McGraw, K. M., & Stroh, P. (1989). An impression-driven model of candidate evaluation. *American Political Science Review, 83,* 399–419.

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

Loftus, E. F. (2005) Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning and Memory, 12,* 361–366.

Loken, B., & Hoverstad, R. (1985). Relationships between information recall and subsequent attitudes: Some exploratory findings. *Journal of Consumer Research, 12,* 155–168.

McConnell, A. R., Sherman, S. J., & Hamilton, D. L. (1994). On-line and memory-based aspects of individual and group target judgements. *Journal of Personality and Social Psychology, 67,* 173–185.

McConnell, A. R., Sherman, S. J., & Hamilton, D. L. (1997). Target entitativity: Implications for information processing about individual and group targets. *Journal of Personality and Social Psychology, 72,* 750–762.

Minsky, M. (1975). A framework for representing knowledge. In P. H Winston (Ed.), *The psychology of computer vision* (pp. 211–277). New York: McGraw-Hill.

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

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

Murray, D. J. (1966). Vocalization-at-presentation and immediate recall with varying recall methods. *Quarterly Journal of Experimental Psychology, 18,* 9–18.

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition, 18,* 251–269.

Nairne, J. S. (2005). Modeling distinctiveness: Implications for general memory theory. In R. R. Hunt & J. B. Worthen (Eds.), *Distinctiveness and Memory* (pp. 27–46). New York: Oxford University Press.

Neath, I. (1997). Modality, concreteness, and set-size effects in a free-reconstruction of order task. *Memory & Cognition, 25,* 256–263.

Neath, I., Brown, G. D. A., McCormack, T., Chater, N., & Freeman, R. (2006). Distinctiveness models of memory and absolute identification: Evidence for local, not global, effects. *Quarterly Journal of Experimental Psychology, 59,* 121–135.

Neath, I., & Surprenant, A. M. (2003). *Human memory (second edition)*. London: Wadsworth.

Oishi, S., & Sullivan, H. (2006). The predictive value of daily vs. retrospective well-being judgments in relationship stability. *Journal of Experimental Social Psychology, 42,* 460–470.

Oliver, A. (2008). Assessing the influence of Gestalt-type characteristics on preferences over lifetime health profiles. *Medical Decision Making, 28,* 723–731.

Page, L., & Page, K. (in press). Last shall be first: A field study of biases in sequential performance evaluation of the Idol series. *Journal of Economic Behavior and Organization.*

Pennington, N., & Hastie, R. (1986). Evidence evaluation in complex decision making. *Journal of Personality and Social Psychology, 51,* 242–258.

Pennington, N., & Hastie, R. (1992). Explaining the evidence: Tests of the Story Model for juror decision making. *Journal of Personality and Social Psychology, 62,* 189–206.

Rabinowitz, F. M., & Andrews, S. R. (1973). Intentional and incidental learning in children and the von Restorff effect. *Journal of Experimental Psychology, 100,* 315–318.

Read, D. (2004). Utility theory from Jeremy Bentham to Daniel Kahneman. *Economy, 58,* 373–396.

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.

Redelmeier, D. A., Katz, J., & Kahneman, D. (2003). Memories of colonoscopy: a randomized trial. *Pain, 104,* 187–194.

Reyes, R. M, Thompson, W. C., & Bower, G. H. (1980). Judgmental biases resulting from differing availabilities of arguments. *Journal of Personality and Social Psychology, 39,* 2–12.

Reyna, V. F., & Brainerd, C. J. (1992). A fuzzy-trace theory of reasoning and remembering: Pattern, paradoxes, and parallelism. In A. F. Healy, R. M. Shiffrin, & S. M. Kosslyn, (Eds.), *From learning processes to cognitive processes: Essays in honor of William K. Estes* (Vol. 2, pp. 235–260). Hillsdale, NJ: Erlbaum.

Reyna, V. F., & Kiernan, B. (1994). Development of gist versus verbatim memory in sentence recognition: Effects of lexical familiarity, semantic content, encoding instructions, and retention interval. *Developmental Psychology, 30,* 178–191.

Rode, E., Rozin, P., & Durlach, P. (2007). Experienced and remembered pleasure for meals: Duration neglect but minimal, peak, end (recency) or primacy effects. *Appetite, 49,* 18–29.

Rozin, A., Rozin, P., & Goldberg, E. (2004). The feeling of music past: How listeners remember musical affect. *Music perception, 22,* 15–39.

Rubin, D. (1986). *Autobiographical memory.* Cambridge, England: Cambridge University Press.

Rundus, D. (1971). Analysis of rehearsal processes in free recall. *Journal of Experimental Psychology, 89,* 63–77.

Russo, J. E., Carlson, K. A., & Meloy, M. G. (2006). Choosing an inferior alternative. *Psychological Science, 17,* 899–904.

Russo, J. E., Medvec, V. H., & Meloy, M. G. (1996). The distortion of information during decisions. *Organizational Behavior and Human Decision Processes, 66,* 102–110.

Russo, J. E., Meloy, M. G., & Medvec, V. H. (1998). Predecisional distortion of product information. *Journal of Marketing Research, 35,* 438–452.

Russo, J. E., Meloy, M. G., & Wilks, T. J. (2000). Predecisional distortion of information by auditors and salespersons. *Management Science, 46,* 13–27.

Samuelson, P. (1937). A note on measurement of utility. *Review of Economic Studies, 4,* 155–161.

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.

Schreiber, C. A., & Kahneman, D. (2000). Determinants of the remembered utility of aversive sounds. *Journal of Experimental Psychology: General, 129,* 27–42.

Schwarz, N. (1995). Social cognition: Information accessibility and use in social judgment. In E. E. Smith & D. N. Osherson (Eds.), *Thinking. An invitation to Cognitive Science (2nd edition)* (pp.345–376). London: MIT Press.

Schwarz, N. (1998). Accessible content and accessibility experiences: The interplay of declarative and experiential information in judgment. *Personality and Social Psychology Review, 2*, 87–99.

Schwarz, N., Bless, H., Wanke, M. & Winkielman., P. (2003). Accessibility revisited. In G. V. Bodenhausen, & A. J. Lambert (Eds.), *Foundations of Social Cognition: A Festschrift in Honor of Robert S. Wyer, Jr.* (pp. 51–78). Mahwah, N.J: Lawrence Erlbaum Associates.

Schwarz, N., Strack, F., & Mai, H. P. (1991). Assimilation and contrast effects in part–whole question sequences: A conversational logic analysis. *Public Opinion Quarterly, 55,* 3–23.

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.

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

Sprenger, J., & Dougherty, M. R. (2006). Differences between probability and frequency judgments: The role of individual differences in working memory capacity. *Organizational Behavior and Human Decision Processes, 99,* 202–211.

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

Stone, A. A., Broderick, J. E., Kaell, A. T., DelesPaul, P. A. E. G., & Porter, L. E. (2000). Does the Peak-End phenomenon observed in laboratory studies apply to real-world pain in rheumatoid arthritics? *The Journal of Pain, 1,* 212–217.

Stone, A. A., Schwartz, J. E., Broderick, J. E., & Shiffman, S. S. (2005). Variability of momentary pain predicts recall of weekly pain: A consequence of the peak (or salience) memory heuristic. *Personality and Social Psychology Bulletin, 31,* 1340–1346.

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.

Taylor, S. E. (1982). The availability bias in social perception and interaction. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristic and biases* (pp. 190–200). Cambridge, England: Cambridge University Press.

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

Tybout, A. M., Sternthal, B., Malaviya, P., Bakamitsos, G. A., & Park, S. (2005). Information accessibility as a moderator of judgments: The role of content versus retrieval ease. *Journal of Consumer Research, 32,* 76–85.

Varey, C., & Kahneman, D. (1992). Experiences extended across time: Evaluation of moment and episodes. *Journal of Behavioral Decision Making, 5,* 169–185.

Västfjäll, D. (2004). The "end effect" in retrospective sound quality evaluation. *Acoustic Science and Technology, 25,* 170–172.

Weber, E. U., Goldstein, W. M., & Barlas, S. (1995). And let us not forget memory: The role of memory processes and techniques in the study of judgment and choice. In J. Busemeyer & D. L. Medin (Eds.), *The psychology of learning and motivation, volume 32* (pp. 33–81). London: Academic Press.

Weber, E. U., Goldstein, W. M., & Busemeyer, J. R. (1991). Beyond strategies: Implications of memory representation and memory processes for models of judgment and decision making. In W. E. Hockley & S. Lewandowsky (Eds.), *Relating theory and data: Essays on human memory in honor of Bennet B. Murdock* (pp. 75–100). Hillsdale, NJ: Erlbaum.

Weber, E. U., & Johnson, E. J. (2006). Constructing preferences from memory. In S. Lichtenstein & P. Slovic (Eds.), *The construction of preferences* (pp. 397–410). New York: Cambridge University Press.

Weiner, B. (1985). "Spontaneous" causal thinking. *Psychological Bulletin, 97,* 74–84.

Winter, L., & Uleman, J. S. (1984). When are social judgments made? Evidence for the spontaneousness of trait inference. *Journal of Personality and Social Psychology, 47,* 237–252.

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.

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist, 35,* 151–175.

Appendices

Appendix I – Screenshot samples

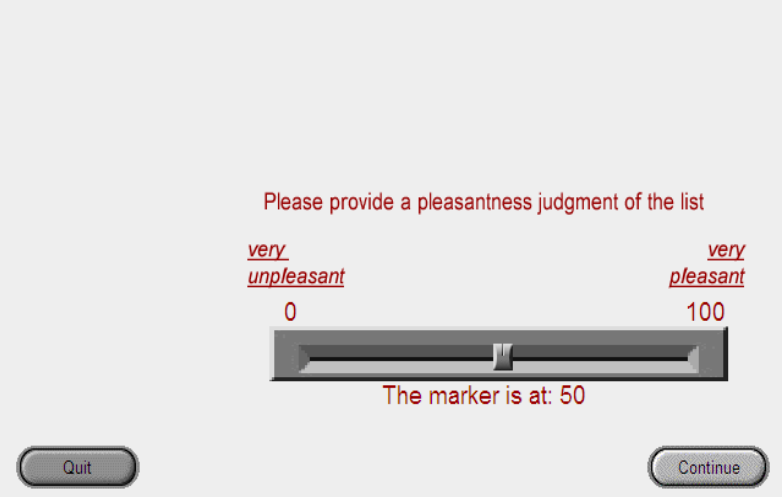


Figure A.1.1. Rating screen, after one participant had initially clicked at the mid-point of the scale (Experiment 2)

Figure A.1.2. One slide from story no. 6 being displayed (Experiments 7–10)

Table A.2.1. *The 42 lists of words used in Experiment 5.*[34] *The asterisks (*) indicate the negative items.*

List no.	Word	VAL	ARO	FAM	KFFRQ	NPHN
1	ulcer*	1.8	6.1	423	5	4
	dove	6.9	3.8	415	4	3
	swamp	5.1	4.9	438	5	5
	statue	5.2	3.5	444	17	6
	garment	6.1	4.5	440	6	6
	sphere	5.3	3.9	457	22	3
	nursery	5.7	4.0	461	13	5
2	deceit*	2.9	5.7	440	2	5
	rattle	5.0	4.4	448	5	4
	hairpin	5.3	3.3	441	1	5
	errand	4.6	3.9	441	7	5
	lantern	5.6	4.1	441	13	6
	context	5.2	4.2	460	35	8
	foam	6.1	5.3	462	37	3
3	assault*	2.0	7.5	470	15	5
	patent	5.3	3.5	426	35	6
	cabinet	5.1	3.4	472	17	7
	fabric	5.3	4.1	477	15	6
	alien	5.6	4.9	479	16	5
	tower	5.5	4.0	463	13	3
	saint	6.5	4.5	463	16	4
4	devil*	2.2	6.1	474	25	4
	silk	6.9	3.7	482	12	4
	cliff	4.7	6.3	479	11	4
	vest	5.3	4.0	472	4	4
	owl	5.8	4.0	477	2	2
	violin	5.4	3.5	468	11	6
	virtue	6.2	4.5	469	30	5
5	riot*	3.0	6.4	490	7	4
	priest	6.4	4.4	484	16	5
	trunk	5.1	4.2	485	8	5
	hawk	5.9	4.4	504	14	3
	journal	5.1	4.1	486	42	4
	stool	4.6	4.0	531	8	4
	crown	6.6	4.3	531	19	4

[34] For each item, the following are reported: ratings from the ANEW database (Bradley & Lang, 1999) about item's valence (VAL) and arousal (ARO); item's familiarity (FAM), Kucera-Francis frequency index (KFFRQ) and number of phonemes (NPHN).

List no.	Word	VAL	ARO	FAM	KFFRQ	NPHN
6	debt*	2.2	5.7	494	13	3
	doll	6.1	4.2	503	10	3
	barrel	5.1	3.4	487	24	5
	nun	4.9	2.9	500	2	3
	tidy	6.3	4.0	490	1	4
	fur	4.5	4.2	530	13	2
	runner	5.7	4.8	531	1	4
7	horror*	2.8	7.2	501	17	4
	bench	4.6	3.6	488	35	5
	tender	6.9	4.9	510	11	5
	tank	5.2	4.9	511	12	4
	cannon	4.9	4.7	498	7	5
	lawn	5.2	4.0	534	15	3
	honey	6.7	4.5	533	25	4
8	poison*	2.0	6.1	504	10	4
	highway	5.9	5.2	488	40	4
	dawn	6.2	4.4	507	28	3
	whistle	5.8	4.7	505	4	4
	frog	5.7	4.5	507	1	4
	autumn	6.3	4.5	533	22	4
	vehicle	6.3	4.6	534	35	5
9	jealousy*	2.5	6.4	500	4	6
	intellect	6.8	4.8	507	5	8
	humane	6.9	4.5	508	5	6
	charm	6.8	5.2	514	26	4
	breeze	6.9	4.4	511	14	4
	poster	5.3	3.9	545	4	5
	engine	5.2	4.0	543	50	5
10	agony*	2.4	6.1	509	9	5
	trumpet	5.8	5.0	490	7	7
	appliance	5.1	4.1	493	5	7
	umbrella	5.2	3.7	511	8	7
	poetry	5.9	4.0	512	88	6
	cottage	6.5	3.4	543	19	5
	cork	5.2	3.8	544	9	3
11	punishment*	2.2	5.9	515	21	9
	nonsense	4.6	4.2	522	13	7
	column	5.2	3.6	519	71	5
	humble	5.9	3.7	519	18	5
	rabbit	6.6	4.0	523	11	5
	tune	6.9	4.7	545	10	4
	flag	6.0	4.6	545	16	4

List no.	Word	VAL	ARO	FAM	KFFRQ	NPHN
12	tragedy*	1.8	6.2	521	49	7
	market	5.7	4.1	518	155	5
	hammer	4.9	4.6	515	9	4
	jelly	5.7	3.7	521	3	4
	spray	5.5	4.1	521	16	4
	chin	5.3	3.3	545	27	4
	item	5.3	3.2	545	54	4
13	crisis*	2.7	5.4	521	82	6
	avenue	5.5	4.1	529	46	6
	vision	6.6	4.7	529	56	6
	passage	5.3	4.4	525	49	5
	queen	6.4	4.8	527	41	4
	patient	5.3	4.2	538	86	6
	social	6.9	5.0	497	380	5
14	failure*	1.7	5.0	542	89	5
	movie	6.9	4.9	523	29	4
	village	5.9	4.1	524	72	5
	tennis	6.0	4.6	528	15	5
	wonder	6.0	5.0	530	67	5
	board	4.8	3.4	546	239	3
	corner	4.4	3.9	556	115	4
15	crime*	2.9	5.4	537	34	4
	theory	5.3	4.6	534	129	4
	nurse	6.1	4.8	537	17	3
	locker	5.2	3.4	538	9	4
	pet	6.8	5.1	541	8	3
	infant	7.0	5.1	513	11	6
	cellar	4.3	4.4	467	26	4
16	anger*	2.3	7.6	541	48	4
	moral	6.2	4.5	535	142	5
	coast	6.0	4.6	541	61	4
	ink	5.1	3.8	542	7	3
	ankle	5.3	4.2	543	8	4
	bland	4.1	3.3	436	3	5
	wink	6.9	5.4	521	7	4
17	hatred*	2.0	6.7	544	20	6
	machine	5.1	3.8	549	103	5
	medicine	5.7	4.4	547	30	7
	opinion	6.3	4.9	550	96	7
	salad	5.7	3.8	554	9	5
	muscular	6.8	5.5	527	16	8
	swift	6.5	5.4	507	32	5

List no.	Word	VAL	ARO	FAM	KFFRQ	NPHN
18	disaster*	1.7	6.3	548	26	7
	method	5.6	3.9	556	142	5
	athletics	6.6	6.1	548	9	8
	stomach	4.8	3.9	547	37	6
	daylight	6.8	4.8	547	15	5
	decorate	6.9	5.1	489	2	7
	solemn	4.3	3.6	453	12	5
19	insult*	2.3	6.0	552	7	5
	yellow	5.6	4.4	555	55	4
	bake	6.2	5.1	549	12	3
	dream	6.7	4.5	553	64	4
	bowl	5.3	3.5	557	23	3
	shadow	4.4	4.3	536	36	4
	plain	4.4	3.5	551	48	4
20	rape*	1.3	6.8	555	5	3
	iron	4.9	3.8	555	43	3
	fish	6.0	4.0	548	35	3
	kettle	5.2	3.2	551	3	4
	paint	5.6	4.1	551	37	4
	muddy	4.4	4.1	541	10	4
	trust	6.7	5.3	548	52	5
21	accident*	2.1	6.3	564	33	8
	industry	5.3	4.5	559	171	8
	beverage	6.8	5.2	566	5	7
	contents	4.9	4.3	566	16	7
	scissors	5.1	4.5	559	1	5
	storm	5.0	5.7	555	26	4
	advantage	7.0	4.8	562	73	8
22	injury*	2.5	5.7	568	27	6
	material	5.3	4.1	559	174	8
	quality	6.3	4.5	560	114	7
	metal	5.0	3.8	559	61	4
	candy	6.5	4.6	559	16	5
	news	5.3	5.2	560	102	4
	mountain	6.6	5.5	574	33	6
23	bomb*	2.1	7.2	566	36	3
	horse	5.9	3.9	560	117	3
	quart	5.4	3.6	568	3	3
	coin	6.0	4.3	564	10	3
	elbow	5.1	3.8	564	10	4
	mail	6.9	5.6	554	47	3
	health	6.8	5.1	577	105	4

List no.	Word	VAL	ARO	FAM	KFFRQ	NPHN
24	pain*	2.1	6.5	569	68	4
	farm	5.5	3.9	564	125	3
	garden	6.7	4.4	567	60	5
	flower	6.6	4.0	566	23	4
	event	6.2	5.1	574	81	5
	youth	6.8	5.7	551	82	3
	taste	6.7	5.2	580	59	4
25	fear*	2.8	7.0	569	127	3
	name	5.6	4.3	573	294	3
	space	6.8	5.1	576	184	4
	wine	6.0	4.8	570	72	3
	cook	6.2	4.4	568	47	3
	church	6.3	4.3	560	348	5
	part	5.1	3.8	579	500	3
26	dead*	1.9	5.7	581	174	3
	office	5.2	4.1	566	255	4
	river	6.9	4.5	565	165	4
	square	4.7	3.2	576	143	4
	pie	6.4	4.2	576	14	2
	white	6.5	4.4	590	365	3
	mind	6.7	5.0	591	325	4
27	sick*	1.9	5.0	571	51	3
	lamp	5.4	3.8	578	18	4
	hat	5.5	4.1	580	56	3
	rock	5.6	4.5	583	75	3
	lake	6.8	4.0	583	54	3
	habit	4.1	4.0	583	23	4
	black	5.4	4.6	603	203	4
28	afraid*	2.0	6.7	575	57	5
	corridor	4.9	3.6	579	17	6
	foot	5.0	3.3	583	70	3
	quiet	5.6	2.8	577	76	5
	circle	5.7	3.9	581	60	4
	wife	6.3	4.9	585	228	3
	dark	4.7	4.3	598	185	3
29	hurt*	1.9	5.9	579	37	3
	milk	6.0	3.7	588	49	4
	seat	5.0	3.0	597	54	3
	fork	5.3	4.0	584	14	3
	grass	6.1	4.1	587	53	4
	rain	5.1	3.7	604	70	3
	red	6.4	5.3	607	197	3

List no.	Word	VAL	ARO	FAM	KFFRQ	NPHN
30	mad*	2.4	6.8	590	39	3
	cat	5.7	4.4	582	23	3
	bar	6.4	5.0	592	82	2
	dress	6.4	4.1	588	67	4
	adult	6.5	4.8	590	25	5
	arm	5.3	3.6	608	94	2
	boy	6.3	4.6	606	242	2
31	angry*	2.9	7.2	600	45	5
	pencil	5.2	3.1	598	34	5
	window	5.9	4.0	621	119	5
	clothing	6.5	4.8	614	20	6
	butter	5.3	3.2	615	27	4
	restaurant	6.8	5.4	593	41	7
	street	5.2	3.4	602	244	5
32	terrible*	1.9	6.3	605	45	6
	teacher	5.7	4.1	599	80	5
	serious	5.1	4.0	593	116	6
	clock	5.1	4.0	608	20	4
	truck	5.5	4.8	620	57	4
	building	5.3	3.9	607	160	7
	answer	6.6	5.4	605	152	4
33	cuisine	6.6	4.4	335	1	6
	banner	5.4	3.8	381	8	4
	glacier	5.5	4.2	409	1	6
	kerosene	4.8	4.3	418	6	7
	gender	5.7	4.4	450	2	5
	egg	5.3	3.8	608	12	2
	chair	5.1	3.2	617	66	3
34	nectar	6.9	3.9	344	3	5
	agility	6.5	4.9	397	3	7
	prairie	5.8	3.4	416	21	5
	kerchief	5.1	3.4	439	1	6
	mystic	6.0	4.8	455	3	6
	sugar	6.7	5.6	608	34	4
	tree	6.3	3.4	613	59	3
35	sentiment	6.0	4.4	471	23	9
	thermometer	4.7	3.8	481	10	8
	hay	5.2	4.0	486	19	2
	coarse	4.6	4.2	506	10	3
	unit	5.6	3.8	513	103	5
	glass	4.8	4.3	611	99	4
	letter	6.6	4.9	610	145	4

List no.	Word	VAL	ARO	FAM	KFFRQ	NPHN
36	cord	5.1	3.5	477	6	3
	basket	5.5	3.6	485	17	6
	indifferent	4.6	3.2	502	11	8
	pig	5.1	4.2	509	8	3
	phase	5.2	4.0	516	72	3
	radio	6.7	4.8	644	120	5
	colour	7.0	4.7	582	141	4
37	stove	5.0	4.5	525	15	4
	cow	5.6	3.5	529	29	2
	employment	6.5	5.3	531	47	9
	tool	5.2	4.3	532	40	3
	custom	5.9	4.7	538	14	6
	person	6.3	4.2	620	175	4
	paper	5.2	2.5	635	157	4
38	lamb	5.9	3.4	519	7	3
	art	6.7	4.9	529	208	2
	orchestra	6.0	3.5	533	60	7
	obey	4.5	4.2	538	8	3
	bandage	4.5	3.9	546	4	6
	girl	6.9	4.3	645	220	3
	book	5.7	4.2	643	193	3
39	modest	5.8	4.0	546	29	6
	army	4.7	5.0	555	132	3
	save	6.5	5.0	559	62	3
	moment	5.8	3.8	560	246	6
	hotel	6.0	4.8	565	126	5
	door	5.1	3.8	630	312	2
	face	6.4	5.0	612	371	3
40	manner	5.6	4.6	546	124	4
	ship	5.6	4.4	553	83	3
	chance	6.0	5.4	563	131	5
	history	5.2	3.9	564	286	7
	smooth	6.6	4.9	570	42	4
	woman	6.6	5.3	623	224	5
	air	6.3	4.1	608	257	1
41	green	6.2	4.3	583	116	4
	month	5.2	4.0	598	130	4
	blue	6.8	4.3	593	143	3
	natural	6.6	4.1	594	156	6
	bottle	6.2	4.8	591	76	4
	city	6.0	5.2	616	393	4
	thought	6.4	4.8	603	515	3

List no.	Word	VAL	ARO	FAM	KFFRQ	NPHN
42	nice	6.6	4.4	583	75	3
	table	5.2	2.9	599	198	4
	plant	6.0	3.6	592	125	5
	key	5.7	3.7	603	88	2
	watch	5.8	4.1	576	81	4
	hand	6.0	4.4	601	431	4
	water	6.6	5.0	641	442	4

Figure A.2.1. Story example: pictures and narratives used for story no. 8 (Experiments 7–10)

A. Sophie works as a legal secretary**

B. Sophie goes to the office downtown

C. On her way, she sees a man playing music

D. At home, she looks out on the lake

E. She takes a stool back to the kitchen

F. Sophie thinks of her sister who committed suicide*

* Negative slide

** Neutral slide matched for familiarity with the negative slide

Table A.2.2. *Pictures and narratives used for the 16 slideshow stories (Experiments 7–10).*[35]

Story no.	Picture ID	VAL	ARO	FAM	Narratives
1	5972*	1.6	4.6	4.4	Tony and Jane learn that yesterday's tornados claimed lives
	7006	3.7	2.2	5.7	They start setting the dining room table for dinner
	4605	5.0	3.1	5.2	Tony and Jane do the crosswords in the daily paper
	8260	4.2	3.2	5.1	Tony watches the grand prix motorcycle race on television
	7500	3.9	1.7	5.7	Jane does some work on her latest project as an architect
	2880**	4.0	1.7	4.6	Tony and Jane look at some pictures together
2	9910*	1.8	4.1	4.2	Matt looks out of his window and sees a car accident
	7031	3.8	2.0	6.0	He takes his old shoes off and puts them to one side
	5410**	5.3	3.2	4.5	Matt listens to some music on the radio
	7351	5.2	3.2	5.6	He puts his favourite pizza in the oven
	7217	4.0	1.6	6.1	He gets into more comfortable clothes and hangs-up his jacket
	5740	5.2	1.6	6.7	Matt waters his new plant in the rear garden
3	2205*	1.8	3.7	4.0	Robert thinks of his old parents
	2840**	3.8	2.5	4.0	Robert plays chess with his son
	7150	4.9	2.3	6.4	It starts raining outside
	8311	5.3	2.9	5.9	He watches some golf on TV
	8320	5.0	2.6	4.2	He watches a bit of car racing
	5731	5.8	2.7	5.4	Robert goes out for a walk
4	2141*	1.8	3.2	3.6	Sarah starts to organise her father's funeral
	7050	4.0	1.9	6.0	She washes and dries her hair
	2221**	3.7	2.7	3.6	Sarah calls her paternal uncle
	7700	4.1	2.7	5.4	She starts tidying-up the storage room
	5220	5.1	3.1	5.9	She goes out for a walk in the nearby park
	7620	4.5	2.6	6.2	Sarah picks up her husband at the airport

[35] For each picture, the following are reported: the picture number within the IAPS database (Picture ID; Lang, Bradley, & Cuthbert, 1999), its valence (VAL), arousal (ARO), and familiarity (FAM). The narratives that accompanied each slide are reported, too.

Story no.	Picture ID	VAL	ARO	FAM	Narratives
5	2751*	1.7	5.5	4.9	George gets fined for drinking while driving
	7010	4.4	1.6	4.1	He prepares to go and get the clothes on the line
	7030	1.6	4.2	1.9	He gets ready to iron some clothes
	7550	4.1	1.9	6.7	George works on his desktop computer
	7320	4.9	2.2	5.3	He eats some fresh fruit as a snack
	7496**	4.0	3.0	4.9	George goes out for some groceries
6	2900*	1.9	3.6	5.6	Jonathan learns that his son is bullied at school
	7224	4.3	1.2	7.0	Jonathan has a busy morning at the office
	1600	6.0	3.0	6.4	In the afternoon, he goes for a horse riding lesson
	2749**	4.1	2.1	5.4	Jonathan smokes a cigar and watches some TV
	7590	4.1	3.1	6.4	He drives to a council meeting
	7130	4.2	2.3	7.2	A sofa is delivered to his house
7	2691*	2.0	4.3	2.6	James hears on the news about violent riots
	7080	4.2	2.2	6.7	He gets the cutlery out and prepares some food
	2575**	4.1	2.0	2.5	James works on a new propeller design
	2570	4.6	2.3	6.4	James works as a professional engineer
	7560	4.0	2.8	6.7	He drives down the motorway to get home
	5395	4.6	3.0	5.4	He goes for a short walk in the harbour
8	9000*	2.0	4.8	4.8	Sophie thinks of her sister who committed suicide
	7025	4.1	1.6	6.7	She takes a stool back to the kitchen
	7705	3.9	1.4	5.5	Sophie goes to the office downtown
	2487	4.3	2.4	4.7	On her way, she sees a man playing music
	2383**	4.3	2.8	4.8	Sophie works as a legal secretary
	5390	6.0	2.0	5.4	At home, she looks out on the lake

Story no.	Picture ID	VAL	ARO	FAM	Narratives
9	6210*	2.0	4.4	4.2	Darren is told that his brother got robbed
	7175	4.6	1.7	8.1	He buys a lamp for his living room
	2210	3.5	1.9	5.0	He comes to the decision to shave his beard
	2600	4.8	2.9	5.4	He meets a cousin of his at the pub
	1450**	4.7	1.8	4.6	Darren reads the latest issue of National Geographic
	7710	5.0	3.2	5.5	Darren makes a start on changing the bed
10	2710*	2.3	3.2	3.0	Peter finds out that his brother takes drugs
	7020	4.8	2.3	7.2	Today, it is a particularly hot and humid day
	7950	4.2	1.9	6.9	He gets a tissue to blow his nose
	2280	4.2	2.2	6.2	Today Peter goes for a trip with his family
	7490**	4.0	2.0	3.2	Peter now lives in the countryside
	7140	4.6	2.4	6.9	He sees a large coach parked by the green
11	2120*	2.7	3.6	4.2	Barry's furious boss calls and fires him
	7170	4.5	2.3	6.4	He changes a light bulb in the hall
	2200	3.8	2.2	5.6	Barry surfs the Internet for a while
	7235	4.7	2.0	7.4	He fixes one of the dining chairs
	7491	4.1	2.4	5.3	In the afternoon, he goes out for a short walk
	5455**	4.1	2.6	3.5	Barry books a return ferry ticket on-line
12	2100*	3.3	3.1	4.3	Paul has a violent argument with his son
	5875**	5.4	2.9	4.5	Paul goes for a bike ride on the hills
	2480	3.9	2.4	5.3	Paul looks outside the window
	2372	4.3	2.2	6.0	He has a long chat with his daughter
	7510	5.0	2.6	6.2	He looks at city break brochures
	2560	5.4	3.9	5.1	He has a picnic with his family

Story no.	Picture ID	VAL	ARO	FAM	Narratives
					His evening music class finishes at
13	7211	4.0	1.9	8.5	6pm
	7002	4.5	2.6	6.4	Mark has a quick shower upstairs
	2000	5.1	2.5	6.8	His stepfather comes to pick him up
					Mark gets ready for his day at
	2270	5.3	2.4	6.7	school
					Mark eats some small pancakes for
	7286	5.1	2.8	5.8	breakfast
					They drive back home through the
	7595	4.8	2.0	7.7	city
					He finds his rolling pin in the
14	7000	4.6	1.6	6.4	bottom drawer
					Jack wants to organise a dinner
	2010	5.3	2.1	6.8	with some friends
					Jack gets his old recipe book from
	7090	4.5	1.8	6.3	the shelf
					He prepares some desserts and
	7402	4.9	3.6	5.7	arranges them on a plate
					Jack goes to buy some missing
	7495	5.1	2.5	5.9	ingredients
					The end of the day is warm and
	5720	5.0	2.8	6.4	sunny
					He washes one of his mugs before
15	7009	4.5	1.8	6.3	making tea
	7034	3.8	1.3	7.7	Steven does some DIY in the house
					He calls a colleague of his at lunch
	2215	4.4	2.1	5.6	time
	2020	5.2	1.4	6.2	Steven has a day off from the office
					He watches a documentary on
	1560	4.3	3.3	3.9	wildlife
					He cooks pasta with fresh
	5500	4.1	2.1	6.3	mushrooms
					He prepares to have an afternoon
16	7004	4.1	2.0	7.2	snack
	7035	4.6	2.0	7.0	He drinks one of his usual beers
					Andrew is just back from a trip
	2214	4.5	2.4	5.6	abroad
					He goes to the restaurant with his
	7289	5.0	3.0	6.0	wife
					Andrew washes and irons his
	7234	4.4	2.3	6.3	clothes
	5020	6.0	3.2	6.2	He goes out to buy some flowers

* Negative slide

** Neutral slide matched for familiarity with the negative slide