Hearing the news of the death of Princess Diana and September 11: How special are flashbulb memories?

Lia Kvavilashvili

University of Hertfordshire

ABSTRACT

Some events produce vivid and detailed memories lasting for many years whereas others are less detailed and easily forgotten. What makes some events more memorable than others? What is the role of emotional arousal or trauma in the formation of these vivid memories? Most importantly, if something is remembered very vividly and in considerable detail does this necessarily mean that the memory is veridical? One area of research that has addressed these fundamental questions over the past 25 years is the research on flashbulb memories. These have been defined as particularly vivid and long lasting (autobiographical) memories "for circumstances in which one first learned of a very surprising and consequential (or emotionally arousing) event" (Brown & Kulik, 1977, p. 73). However, a question about the special status of flashbulb memories has remained controversial with some studies showing a good test-retest consistency and others showing substantial distortion and forgetting.

In this talk I will describe a series of studies that examined the flashbulb memories of the death of Princess Diana and the terrorist attack on New York on September 11, 2001. The results of these studies show that the phenomenological characteristics of flashbulb memories show virtually no forgetting. For example, 4- and 7-year old memories of the death of Princess Diana were as detailed, specific and vivid as memories of September 11 assessed after only few days from the attack. The special status and the accuracy of flashbulb memories was examined by investigating the test-retest consistency of flashbulb memories of September 11 and comparing it to the consistency of control memories about hearing some unimportant personal news. The results showed that flashbulb memories are not totally immune to forgetting as the consistency dropped reliably from September 2001 to July/August 2003. However, major distortions occurred in only 8% of cases, and they were not total confabulations. Rather, they referred to another occasion in which one heard of September 11 again later on the same or next day. Most importantly, flashbulb memories were significantly more consistent than memories of the control event even though their test-retest delay was almost twice as long (23-24 months) than for control memories (11-12 months). Taken together, the pattern of results supports the notion that flashbulb memories are special and different from ordinary non-flashbulb memories in terms of mechanisms that may be involved in their encoding and long-term retention. The theoretical and methodological implications of these findings will be discussed.

INTRODUCTION

Flashbulb memories have been defined as particularly vivid, detailed and long lasting autobiographical memories that are accompanied by high levels of confidence in their accuracy. We all have such vivid and detailed memories of personally important, emotionally arousing and/or unusual events from our personal life like for example a car accident we were involved in, the first romantic kiss or the day we failed an important exam, etc. However, in psychological research it has been customary to study these memories by asking people to remember their personal circumstances in which they first heard of the news of very important or tragic public event.

Major public events that have been used for studying flashbulb memories over the past 30 years include the assassination of President John F. Kennedy, the explosion of space shuttle Challenger and the resignation of British Prime Minister Margaret Thatcher. More recent events include the Death of Princess Diana and the terrorist attack on New York. One important point that needs to be stressed in relation to these studies is that participants are not asked to recall the details of the event itself (for example, what exactly happened in New York on 11 September). Instead, they are asked to recall their personal circumstances in which they heard of the news such as where they were, what they were doing or who told them, etc.

For example, in a seminal study of Brown and Kulik (1977), who coined the term "Flashbulb Memories", 99% of participants were able to recall several details about their personal circumstances in which they first heard the news of the assassination of President Kennedy even though 13 years had passed from this event at the time of their testing. The results of this and subsequent studies indicate that remembering personal circumstances of hearing news of public events is a good analogue for studying flashbulb memories of personal events.

In order to explain the long-lasting nature of these memories Brown and Kulik (1977) postulated the existence of special encoding brain mechanism, the so called "print now" mechanism. According to Brown and Kulik (1977) this mechanism switches on when the levels of surprise and importance/consequentiality exceed certain threshold which will then result in a detailed and permanent memory trace.

The aim of the research that I am going to report today was to examine the hypothesis about the special status of flashbulb memories by studying two related issues: the stability of these memories and the accuracy of these memories. The stability concerns the question about how permanent these memories are over long time delays in terms of their phenomenological characteristics irrespective of their accuracy. In other words, will flashbulb memories of important public event that happened many years ago be as specific, detailed and vivid as flashbulb memories of an event that happened only a few days ago? This is an important research question that has not been directly addressed in previous studies (but see Kvavilashvili, Mirani, Schlagman & Kornbrot, 2003).

On the other hand, the issue about accuracy involves asking a different question, namely, if someone has a very vivid, detailed and specific memory of an important flashbulb event does it mean that this memory is accurate? This is a very important question especially in the light of recent debates on the recovered memories of

childhood sexual abuse or the accuracy of eyewitness testimony. In the first half of this talk I will describe the study that examined the issue of stability of flashbulb memories irrespective of their accuracy and in the second part of the talk I will describe a study addressing the issue about the accuracy of flashbulb memories.

In both studies participants were administered the standard flashbulb memory questionnaire which consists of three parts. First, participants are asked to recall their personal circumstances in which they first heard of the news of important public event (i.e., free recall). Next, they are asked five specific questions about what time did they hear the news, where they were, what they were doing, who told them, and whether there were others present. This can be regarded as a probed recall. Finally, participants are asked to provide ratings of their surprise, emotion, importance of event as well as the rating of vividness of their memory image. A 10-point rating scale was used with 1=not at all and 10=extremely.

The phenomenological characteristics of memories can be assessed by the number of details provided in memory descriptions and the specificity of participants' responses to five questions as well as the ratings of vividness. However, the accuracy of memories can only be assessed by administering the Flashbulb Memory Questionnaire twice, and comparing participants' responses given immediately after the event to their responses given after a long delay.

STUDY 1

The aim of Study 1 was to assess the stability of flashbulb memories by examining the following three phenomenological characteristics: The number of details mentioned in memory descriptions, the specificity of responses to 5 questions and the ratings of vividness. The basic idea was to compare these phenomenological characteristics of memories of two different but comparable public events, one with long and another with a shorter delay interval. In fact, we wanted to stretch this comparison to its limit by comparing flashbulb memories of September 11 that were only few days old to the flashbulb memories of the death of Princess Diana that were several years old.

In total, we had 5 different groups of young British participants (see Table 1 below). Two groups of 45 and 39 participants were tested in September 2001 for their flashbulb memories of terrorist attack on New York on September 11, 2001. One group was tested on 12 and 13 September and another group on 20 and 21 September. This means that these flashbulb memories were potentially as detailed, specific and vivid as one can get because the memories were very fresh, only 1-2 days old in Group 1 and 10-11 days old in Group 2. We then tested three additional groups of young British participants at different time points for their memories of the death of Princess Diana. One group was tested in December 2001 at which time their flashbulb memories of Princess Diana's death were 4 years and 3 months old (the fatal car crash happened on 31 August, 1997). The remaining two groups were tested in July/August 2003 and July/August 2004 which means that at the time of the testing their flashbulb memories of the death of Princess Diana were 6 and 7 years old respectively.

Table 1 Information about participants, events, time of testing and age of memories in Study 1

Participants	**Event**	**Time of testing**	**Age of memory**
Group1 (N=45)	September 11	12-13 Sept. 2001	1-2 days
Group2 (N=39)	September 11	20-21 Sept. 2001	10-11 days
Group 3 (N= 65)	Princess Diana	Dec. 2001	4 years
Group 4 (N=89)	Princess Diana	July/Aug. 2003	6 years
Group 5 (N=47)	Princess Diana	July/Aug. 2004	7 years

The results are presented in Table 2 and show the means of the three phenomenological characteristics of memory as a function of group. The mean ratings of vividness of memory image were made on a 10-point scale. The end points of this scale ranged from 1=very vague, almost no image at all to 10=extremely vivid, almost as normal vision. It is important that participants rated the vividness of image of their personal circumstances and not the images of actual events broadcasted by television. As one can see there was no group effects ($F<1$) so that 7-year old memories of the death of Princess Diana were rated as vivid as only 1-2 days old memories of September 11. There was also no statistically significant difference between the groups in the mean number of details provided by participants in their memory descriptions ($F<1$). Finally, we looked at the specificity of participants' responses to 5 questions about the time, location, activity, source and others present. For each question the specificity score could vary from 0 to 2 depending on the specificity of the response. The total specificity score could vary from 0 to 10. As one can see from Table 2, the mean specificity scores are generally very high in all four groups with a slight drop in Group 5. One way between group ANOVA revealed a significant main effect of groups. However, the follow up analysis showed that this effect was due to a significant difference between Group 1 and Group 5 only. Thus, one can conclude that there is very little drop in the specificity of memories at least within the first 6 years of the event.

Table 2 Mean ratings of vividness, number of details in memory descriptions and specificity scores as a function of group. The results of 1-way between subject ANOVAs are presented in the final column.

	Group 1	Group 2	Group 3	Group 4	Group 5	F (4,279)
Vividness	8.13	7.85	7.75	7.84	7.53	$F<1$
No of Details	3.44	3.95	3.49	3.74	3.83	$F<1$
Specificity	9.09	8.72	8.55	8.84	7.79	$F=4.55$

In conclusion, the results of Study 1 provide convincing evidence in support of the idea that there is very little forgetting in flashbulb memories in terms of

phenomenological characteristics of memories. Thus, 6-7year old memories of Princess Diana were as detailed specific and vivid as memories of September 11 that were only few days old. However, the next important question that we wanted to address was the question about the accuracy of these memories.

STUDY 2

As pointed out earlier, the accuracy or the consistency of flashbulb memories can be studied with the test-retest method where participants fill in the flashbulb memory questionnaire twice, once immediately or very soon after the public event and second time after a considerable delay. This allows to assess the consistency of participants' responses at the initial test and at the re-test.

The drop out rate in the test-retest studies of flashbulb memories is notoriously high. There are only two published studies with a delay interval longer than two years and, interestingly, both studies report significant distortions and forgetting in flashbulb memories. In Study 2, we therefore wanted to assess the test-retest consistency of flashbulb memories by using a long delay interval of 2 years and such an important and tragic public event as terrorist attack on New York on September 11, 2001. The enormous impact that this event had on international community gave us a unique opportunity to assess the special status of flashbulb memories. Indeed, if the terrorist attack on New York can not produce long lasting and consistent memories then the hypothesis of special status of flashbulb memories will be seriously compromised.

However, in order to properly assess this hypothesis one needs to compare the consistency of flashbulb memories to that of non-flashbulb control memories. This raises an issue as to what can count as an appropriate control event. Brewer (1992) has suggested that the most appropriate control will be asking participants to remember circumstances in which they heard of some personal but unimportant news, for example hearing that you did not win a prize in the competition. Therefore, a unique feature of the present study was that we staged a control event for the participants and then assessed the test-retest consistency in the same way as for the flashbulb event.

Thus, in the first study (Study 2a) young British participants were administered the Flashbulb Memory Questionnaire in September 2001. Half of the participants were tested on 12 and 13 September and half were tested on 20 and 21 September. Then all participants were re-tested in July/August 2003 after a delay of 23/24 months.

In the second study (Study 2b), a new sample of young participants was recruited in summer 2003 to take part in a study of memories of personally experienced events. They were warned that there would be several telephone interviews and that by taking part they had a chance to win £100 in a prize draw run by the researchers at the beginning of the study. After an initial introductory interview they were warned that the next interview would take place some time soon and were asked to keep their mobile phones switched. In few days, they indeed received a call from a different researcher who told them that the interview was cancelled and additionally informed them that they did not win the prize in the prize draw. This was the staged control event and in subsequent interviews participants were administered the Flashbulb Memory questionnaire and asked to remember their personal circumstance in which they first heard of this unimportant personal news. Like in the flashbulb study (Study

2a), half of the participants were tested 1 or 2 days after hearing the news of not winning the prize and the other half was tested after 10-11 days. All participants were re-tested in summer 2004, 12 months after the event. Unfortunately, it was not possible to have a delay of 24 months for the control event. However, we reasoned that if flashbulb memories are indeed special and different from control memories then their test-retest consistency should still be significantly better after almost twice as long delay.

In Study 2a, there were 84 young participants (mean age 32 years, range 20-56) and in Study 2b – 89 young participant (mean age 37 years, range 22-50). The drop out rate after 24 and 12 months respectively was remarkably low so that the final number of participants who were tested on both occasions was 64 and 79 participants in Study 2a and 2b, respectively.

Having conducted the study the next stage was to code participants' responses on the Flashbulb Memory Questionnaire for the consistency. This was done separately for participants' answers to 5 specific questions in the probed recall and to participants' memory descriptions in free recall. When coding for the consistency of 5 questions we used the scoring system devised by Neisser and Harsh (1992). In this system, participants' responses to each question at the re-test are compared to their responses at an initial test and assigned a score of 0 to 2 depending on the match between the two responses. For example, a score of '0' is assigned if, at the re-test, participant does not remember where he was or provides a completely different location to the Place question. A score of '1' is assigned if the participant's response is less specific or only slightly incorrect (for example if initially the participant said at home in my bedroom but at the re-test they say at home or at home but in the kitchen). A score of '2' is assigned if the participant's response is exactly the same or even more specific than at an initial test. The so called Weighted Attribute Score proposed by Neisser and Harsh (1992) puts particular emphasis on Place, Activity and Source and regards Time and Others Present as less important attributes of flashbulb memories. The scoring system takes this into account by assigning the maximum of 2 points for Place, Activity and Source and only one extra point if participant obtains a score of 3 or above for the Time and Others present. The resultant Weighted Attribute Score varies from 0 to 7 with higher scores indicating higher levels of consistency.

The mean Weighted Attribute Scores were entered into a 2 by 2 between subjects ANOVA with the event and delay interval as independent variables. The delay interval (i.e., whether participants were initially tested after 1/2 or 10/11 days) did not have any effect on memory consistency. However, there was a highly significant effect of event ($F(1,284)=91.50$, $p<.0001$, effect size-partial eta squared=.26). The consistency scores were reliably higher for flashbulb memories of September 11 ($M=5.13$, $SD=1.60$) than for the control event ($M=3.20$, $SD=1.73$) even though the test-retest interval was 24 months for September 11 and only 12 months for the control event. Importantly, the consistency score for September 11 is also reliably higher than the consistency scores reported in two previous flashbulb memory studies with long test-retest delays (2.95 in the study of Neisser & Harsh, 1992, and 3.30 in the study of Scmolck, Buffalo, & Squire, 2000).

However, one problem with the Weighted Attribute scores is that they do not distinguish the 'don't remember' response from major distortions. Consider the

following memory description provided by one male participant on 21 September, 2001: "I was returning from a shopping trip and I put on the car radio and heard something on the news and it was very early and the news story was just breaking and they interrupted Radio 4… As I arrived home I just sat and watched it". Now compare it to this description provided in July 2003, two years after the initial test: "We were standing in a queue (Jenni and I) waiting to, check in for a flight to Nimes the day after – the 12th of September (a Wednesday or Thursday). We were quite horrorstruck and surprised. We had been packing up in the night before and we were subject when we checked in to the most vigorous and thorough luggage and body search to Nimes".

The problem with Weighted Attribute Score is that this participant will get a score of 0 together with the participants who acknowledge at the re-test that they simply do not remember. In order to take this distinction into account and to get a better idea about what was really happening with participants' memories we separately examined participants' memory descriptions and coded them into the following six categories:

Don't remember
Major distortion (as in the above case)
Minor distortion (if the memory description was mostly correct but one or two details were incorrect)
Less specific (if the description was same but contained less details)
More specific (if the description was the same plus contained additional details)
Same (if the description at re-test was exactly the same as at initial test).

Table 3 presents the percentage of participants whose memories were classed into these six categories as a function of event: flashbulb versus control. The chi-square test on this data revealed a highly significant effect of event (χ^2=62.40, N=143, df=5, p<.0001). However, the table shows that the difference is not in the categories of major or minor distortions. The number of memories classed into these categories are surprisingly low for both types of events. The major difference between the two events lies in the categories of "Less specific" and "More specific". While 62% of control memories became less specific after 12 months only 14% of flashbulb memories became so after 24 months. Moreover, while 44% of flashbulb memories became more specific only 1% of control memories became more specific.

Table 3 The percentages (raw numbers in brackets) of participants whose memories were classed into six categories as a function of event (flashbulb vs. control)

	CONSISTENCY OF MEMORY DESCRIPTIONS						
	Don't Remember	Major Distortion	Minor Distortion	Less Specific	More Specific	Same	Total
Flashbulb	0% (0)	8% (5)	22% (14)	14% (9)	44% (28)	12% (8)	100% (64)
Control	9% (7)	11.5% (9)	11.5% (9)	62% (49)	1% (1)	5% (4)	100% (79)

In conclusion, several interesting findings emerged from this study. First finding is that the consistency was markedly better than in previous two studies with long delays by Neisser and Harsh (1992) and Schmolck et al. (2000). This superior consistency was obvious both in terms of Weighted Attribute Scores and especially in terms of the percentage of memory descriptions classed as major distortions. Thus, in the study of Neisser & Harsh (1992), 25% of memories were classed as major distortions and in the study of Schmolck et al (2000) this percentage was as high as 40%. In contrast, in our study only 8% of memories were classed as major distortions.

However, the mere fact that there was 8% of major distortions and 22% of minor distortions indicates that flashbulb memories are not totally immune to forgetting as originally suggested by Brown and Kulik (1992). This is also obvious from the mean Weighted Attribute Score for September 11 that it was not at ceiling, the mean score was 5.13 out of possible 7 after two years from the event.

However, by far the most important and novel finding that emerged from the study is that the retention rate was significantly better for flashbulb memories than for memories of the staged control event. If we had not compared flashbulb memories to control memories we would have concluded that flashbulb memories are not special and different from other autobiographical memories of more ordinary events. However, the results clearly show that flashbulb memories are retained in quantitatively and qualitatively different ways than non-flashbulb memories.

CONCLUSIONS

The overall conclusion, based on the results of our studies, is that flashbulb memories are indeed special and different from ordinary autobiographical memories. Although they may not be totally immune to forgetting they do show remarkable permanence in phenomenological characteristics over as many as 6-7 years as shown in Study 1 and significantly better consistency than memories of control events as shown in Study 2.

The present findings contrast the currently prevailing view held by flashbulb memory researchers who claim that flashbulb memories are prone to distortions and forgetting like ordinary, non-flashbulb memories. For example, according Talarico and Rubin (2003) the only difference between the two is that flashbulb memories are accompanied by high levels of confidence in their accuracy but are no more accurate than ordinary memories. However, the results of our study show that flashbulb memories are far more consistent than ordinary non-flashbulb memories.

In fact, I believe that some distortions in flashbulb memories found in our study are due to research methodology and that flashbulb memories of personal events are even more accurate than the ones studied via important public event. The problem with public events is that they are constantly televised and talked about by everyone so that each person would hear the news again many times during the day from different sources. Therefore, over the time it may become difficult to distinguish between different memories of hearing the news and to know which one is when you heard the news very first time. In other words, the memory distortions are not complete confabulations but they represent a real but different occasion in which the participants heard about the news. For example, there is no doubt that the participant whose memory descriptions I showed you earlier did indeed hear the news again on the next day at the airport on his way to Nimes. However, he incorrectly remembered

this as the first time he heard this news. This would also explain why people are so confident in their memories classed as "major distortion". Therefore, my prediction is that the percentage of distortions will be much lower if it was possible to study the consistency of flashbulb memories via personal events rather than public events.

It is also interesting that recent studies on the brain mechanisms involved in the formation of emotional and especially traumatic memories provide further support for the idea that flashbulb memories are special. In particular, studies by Cahill and his colleagues have shown the involvement of noradrenergic system and amygdala in the formation of emotional memories. For example, several studies have shown that the administration of beta-adrenergic receptor antagonist propranolol selectively impairs memory for emotionally salient film episode but does not affect the memory for non-emotional episodes (Cahill & McGaugh, 1998). O'Caroll and colleagues have also pointed out the adaptive value of such brain mechanism which helps the organism to better recall potentially dangerous situations (O'Caroll et al., 1999).

Finally, I wish to acknowledge the support of Economic and Social Research Council in conducting research on flashbulb memories and the contribution of my collaborator Diana Kornbrot as well as numerous research assistants who helped me to collect and code the data. Also, to finish on a lighter note, it appears that even animals have flashbulb memories as portrayed by this cartoon. Apparently all forest animals can remember to this day exactly where they were and what they were doing when they first heard about the news that Bamby's mother was shot by an evil hunter.

REFERENCES

Brewer, W. F. (1992). The theoretical and empirical status of the flashbulb memory hypothesis. In E. Winograd & U. Neisser (Eds.), *Affect and accuracy in recall: Studies of "flashbulb" memories* (pp. 274-305). Cambridge: Cambridge University Press.

Brown, R., & Kulik, J. (1977). Flashbulb memories. *Cognition, 5*, 73-99.

Cahill, L., & McGaugh, J. L. (1998). Mechanisms of emotional arousal and lasting declarative memory. *Trends in Neurosciences, 21*, 294-299.

Kvavilshvili, L., Mirani, J., Schlagman, S., & Kornbrot, D. E. (2003). Comparing flashbulb memories of September 11 and the death of Princess Diana: Effects of time delays and nationality. *Applied Cognitive Psychology, 17*, 1017-1031.

Neisser, U., & Harsh, N. (1992). Phantom flashbulbs: False recollections of hearing the news about Challenger. In E. Winograd, & U. Neisser (Eds.), *Affect and accuracy in recall: Studies of "flashbulb memories"* (pp. 9-31). Cambridge: Cambridge University Press.

O'Carroll, R. E., Drysdale, E., Cahill, L., Shajahan, P., & Ebmeier, K.P. (1999). Stimulation of the noradrenergic system enhances and blockade reduces memory for emotional material in man. *Psychological Medicine, 29*, 1083-1088.

Talarico, J. M., & Rubin, D. C. (2003). Confidence, not consistency characterizes flashbulb memories. *Psychological Science, 14*, 455-461.

Schmolck, H., Buffalo, E. A., & Squire, L. R. (2000). Memory distortions develop over time: Recollections of the O.J. Simpson's trial verdict after 15 and 32 months. *Psychological Science, 11*, 39-45.