CHAPTER 2: FLASHBULB MEMORY METHODS[1]

Daniel B. Wright
Florida International University

The observation that some events are more memorable than others, and that some of the events on the high side of the memorability distribution are important news events, is the empirical basis of the flashbulb memory (FBM) phenomenon. As a scientific finding, it is trivial. Of course events vary in how memorable they are (and hence have a distribution), and no prizes are won for saying that some of the events that occur throughout this distribution, including the most memorable end, are news events. There are some interesting scientific questions that can be asked about FBMs, but these questions must go beyond this initial observation. Many FBM studies fail to address the more interesting questions because the way the initial FBM research (Brown & Kulik, 1977) was carried out prompted many subsequent researchers to design studies that are not appropriate to answer many of the interesting questions. Many (not all) of the scientific questions of interest to cognitive researchers about FBMs are causal, while the methods used in the prototypical FBM study are more appropriate for addressing associative questions. Further, the methods that tend to be used are well suited for differentiating among people and groups, which is arguably of more interest to sociologists, oral historians, and social psychologists, than to cognitive psychologists, who are usually more interested in differences between situations and tasks (Wright & Gaskell, 1995).

Before progressing, it is worth making two things clear. First, when referring to the typical FBM study I am referring to a study where a hundred or so people from an opportunity sample are asked questions about what they were doing when they heard about a single news event. Second, the criticisms I raise are partially self-directed (for example, Wright, 1993), so the more pithy remarks are about me. I am playing devil's advocate, to some extent, but this approach highlights the difficulties.

The two basic requirements for the scientific process are measurement and theory. You must have measurement and some measurement theory (Hand, 2004) and explore some set of hypotheses that are part of complex network of hypotheses (Wright, 2006). This chapter is divided into two sections on the basis of these two requirements. The first section examines what FBMs are and what they are not, and how researchers can decide whether a memory is a FBM. I describe various choices researchers have to make in deciding what their domain of study is and some of the methods that could be used. The second section focuses on the main scientific questions that can be asked once the conceptual issues of the first section have been addressed.

I will not discuss details of the different FBM models because there are excellent reviews in other parts of this book (for example, Luminet, Chapter 3) and also a major theme of this chapter is that there are still matters about FBM that need to be resolved before details discussion of the models should be made. However, it is beneficial to produce a simplified schematic of the basic model (Figure 2.1). Most FBM researchers describe how some combination of emotion, importance, and survival impact combine so that events which score high on this combination evoke some special mechanism which produces good memories. How these event characteristics combine is where the detailed discussions of FBM theorizing lay. The special mechanism hypothesis is that the qualities of FBMs cannot be accounted for by the variability in memories produced by normal processes.

Insert Figure 2.1 about here

CONCEPTUAL AND MEASUREMENT QUESTIONS ABOUT FLASHBULB MEMORIES

What is the population under investigation?

Figure 2.2 shows a distribution of memorability for a set of events. For now, assume that "memorability" is some variable about the quality of memories and that beyond some value these are FBMs and if the value is less the memory is not a FBM. People have used different ways to conceptualise this and to measure it, and these are discussed below. As with any distribution, before asking any empirical questions, it is necessary to ask "what is the population of events that form this distribution?" It is unlikely that anyone believes that the population of interest is just FBMs. If they did they would encounter conceptual problems categorizing FBMs and theoretical problems trying to describe what FBMs are without saying what they are not (i.e., discriminate validity). So what is the population? Is it all events experienced? All news events? All events that people recall vividly? All assassinations? Similarly, is the distribution for all people, some selected sub-population, or just an individual?

FIGURE 2.2 NEAR HERE

Brown and Kulik's (1977) methods provide clues of their intent. While most of the events that they asked their sample about were news events, they also asked people to report about personal events. Thus, they were probably interested in all events that might be recalled vividly. Rather than probing Brown and Kulik for the population of events that their catchy phrase, flashbulb memories, was referring to some researchers (e.g., McCloskey, Wible & Cohen, 1988; Neisser, 1982) attacked their assumption that FBMs were accurate (how to differentiate the good and less good memories from Figure 2.1), and others (e.g., Conway et al., 1994; Finkenauer et al., 1998) went onto to develop Brown and Kulik's FBM model (how the emotion, importance, and survival qualities from Figure 2.1 combine into memorability). Unfortunately, Brown and Kulik presented essential no data in their original study to support their model[2], and it is unclear how much further insight subsequent research has given to the concept. While memory accuracy and models of memory are clearly important, making statements about memories for events in some part of Figure 2.2 without considering the population being examined makes inference difficult.

It is worth thinking about some of the possible populations of events. This is a difficult task, since there are not agreed upon definitions of what is an event or what is a memory. There are questions about how microscopic or extended an event should be, and what relationship the person needs to have with the event. These are problems in general for memory research (Rubin, 1992). One possibility is considering all events which at the time somebody with a typical adult's cognitive abilities would be able to narrate their experiences of. This set of events is what many autobiographical memory researchers might consider their domain of interest. Another possibility is all events that can, given the right prompt, be recalled at testing. One of the differences between these two approaches is that many of the events that occurred will not be remembered, so the former would include events low in memorability (below some threshold for recall) and the latter would not. The latter, however, would make it easier to study false memories.

Consider two possible populations. The first is the population defined above and often assumed, though not always explicitly, for much autobiographical memory research: all events for which a typical adult could have given a reasonable account at the time of the event. This allows the population under investigation to include events that some people would forget. This is important because forgetting events, including traumatic events, is an important area of interest (Brewin, 2003; McNally, 2003). This population has the disadvantage that it would not include events that did not happen, so this population would not be appropriate for researching false memories[3].

The second population is one where the researchers are interested in a single event. For example, later in the chapter Judith Zur's (1998) research on La Violencia is described. Her interest was primarily aimed to document this event, not to evaluate different memory theories. Psychologists often use case studies either to illustrate a theory or to falsify one. A famous example from psychology is when Festinger, Riecken and Schacter (1956) used the observation of a single event to illustrate cognitive dissonance. When a group of people, who believed that at a specific time aliens from the planet Clarion would save them from a great flood, realized this was not going to happen, their beliefs became stronger, in line with cognitive dissonance theory. An example from FBM research is using the incredibly vivid personal memories astronomers at the Anglo-Australian observatory had at the discovery of Supernova 1987-A (Wright & Gaskell, 1995). This illustrated that strong personal feelings can be associated with FBM memories for an event of relatively little societal importance.

What should the population of events/memories be for cognitive psychologists?
Several possibilities. One good population is all events that if experienced could be described by the typical adult.

What is Memorability and how to measure it?

The next question is: what is the x axis in Figure 2.2? If it is used to classify FBMs, then it must be something for which FBMs are at one extreme. Brown and Kulik (1977) and subsequent FBM researchers suggest that the prototypical FBM is an extremely long-lasting and vivid recollection. I will consider three ways of conceptualising memorability (see also Curci, Chapter 1). The first two (categorical ratings based on Tulving's (1985) remember/know memory distinction and clarity ratings) allow the participants to define memorability, although for the first it is a discrete measure and the second a continuous measure. The final way uses the responses to memory questions and creates latent variables out of these.

Tulving (1985) differentiated two types of memories. He described remember memories as those which allow the person to mentally re-live the event. He described how they evoked autonoetic consciousness, a way to place oneself in the memory. He described know memories as those which lacked the re-living aspect of remember memories. This distinction is used in many memory studies (see also Gardiner, 1988) where participants are shown, for example, a list of words and later asked if they have seen these words before. If they say "yes" they are asked to write R if it is a remember memory and K if it is a know memory. Within autobiographical memory research there has been a move away from having participants choose between just a K (for Know) and an R (for Remember) response to choosing among K, R, and F or G (for Familiar or

Guess) responses because researchers felt that some participants were using K responses inappropriately (Gardiner, Ramponi & Richardson-Klavehn, 2002).

The re-living self-referential descriptions of remember memories suggests these might be similar to what researchers mean when they describe FBMs. The problem for FBM research is that in practice R responses do not capture what people usually think of as a true FBM. In a typical recognition memory study people may be shown 50 words and later asked if their memory is a K or an R memory. Many respond R, but it is questionable whether these memories are FBMs. What would be necessary is an additional category, like SRM (for super-remember-memory) to capture what it is to be a FBM. The problem with this is Tulving's original distinction and the addition of the F category are both based on research from different areas that suggest these are qualitatively different types of memory. More research would be necessary to see if the F/K/R/SRM taxonomy was appropriate.

If four ordered categories are used, it is likely that many people would interpret the choice in a similar manner to a rating scale from "not a clear memory" to "a very clear memory". This is another alternative: ask people directly for their ratings on a clarity scale. This assumes that the scale is unidimensional and it relies on people's individual interpretations of the scale being similar. Another difficulty is that asking only a single question can be unreliable. Finally, many researchers believe that "clarity" is not the only aspect that should contribute to the flashbulb-ness dimension, but that consistency, details of memory, etc., should also be used in the measurement.

The third approach circumvents the problem of having only a single question. Brown and Kulik (1977) describe one of the defining characteristics of a FBM that people can recall the four Ws: what they were doing, how they heard, who they were with, and where they were[4]. Responses to these questions can be aggregated in some way to define the x axis. Three ways of aggregating the data are considered here: counting the number of responses, conducting a latent trait model, and conducting a latent class model. Moustaki (1996) provides detailed comparisons among these three methods and she does this with specific reference to FBMs so her paper is particularly valuable for FBM researchers (see also Curci, Chapter 1). Here these methods are briefly described.

The "counting the number of responses" approach involves either counting the number of times someone positively responds to each of 4 Ws or counting them but allowing some partial credit. The latent trait and latent class models deserve further explanation. Bartholomew and Knott (1999) described how several different latent variable models all have a similar form, and actually use FBM data to aid their explanation. Latent trait models take several categorical observed variables and try to account for covariation in responses by assuming that there are a small number (often one) of continuous latent variables underlying the responses. They are popular in education, where they are often called item response models, and are analogous to exploratory factor analysis when the observed variables are categorical. This procedure would create a dimension for the x axis of Figure 2.2. Latent class models account for the responses by assuming that there are groups of people. People within the same group have similar probabilities for responding in certain ways, but there are large differences between people of different groups. These are popular in sociology and are becoming more popular in psychology. As discussed in the next section, it is sometimes difficult to differentiate these approaches on an empirical basis.

Are there differences between these approaches?

The choice of what the x axis in Figure 2.2 is and how to measure it are important and inter-related questions. I will compare a couple of the choices for illustrative purposes. First, consider whether to use participants' responses to a clarity scale or the number of positive responses to the what, how, where, and who questions. In Wright et al. (1998, Exp. 2) a large number of participants from a representative sample of the UK adult population were asked about Margaret Thatcher's resignation. They were asked how clear their memory was of the event on a 1 (Cannot remember the event) to 5 (Completely clear) rating scale, and also asked for their responses on the four W-questions. Given that presumably psychologists are astounded by people being able to recall these Ws, otherwise why would Cognition have published Brown and Kulik's original paper, one would expect that people who could recall all four Ws would give themselves very high ratings on the clarity scale. However, people who responded positively to all of these gave on average responses near the midpoint of the clarity scale, certainly not responses that are commensurate with the live, photographic reliving descriptions often given to FBMs. Thus, the choice of using responses to the Ws or rating scales can provide very different interpretations for what FBMs are and produce very different estimates for the number of people with FBMs for any given event.

The second comparison is between latent trait and latent class analysis. Moustaki (1996; see also Curci, 2005, Chapter 1) goes through the Thatcher resignation data in detail. She shows how the data can be used in these models to produce either a single latent trait along which people vary or two latent classes to which people can belong. Bartholomew and Knott (1999; for an introductory textbook on these procedures, see Bartholomew et al., 2002) discuss how it is very difficult to differentiate these models empirically and that the choice often depends on the particular theory one has for FBMs. For example, there is much discussion about whether memory retrieval is based on a threshold model, where people either remember the event or not, versus whether there is some continuous memory strength dimension that predicts the probability of remembering the event. If you assume that people either have or do not have a FBM, as if some trigger either fires or does not fire, then a latent class model is probably more appropriate. However, if you assume that FBMs are simply at one end of some memorability scale then the latent trait model is probably more appropriate. Brown and Kulik's original description of FBMs, where they adopt Livingston's (1967) "Now Print" mechanism, suggests a threshold model which is more consistent with the latent class models. Thus, the class versus trait question is of importance to the "special mechanism" debate if one believes the mechanism is like what Brown and Kulik described (and that ordinary memories vary along a continuum). However, other types of "special mechanisms" could be more continuous in nature, and therefore the choice of class versus trait does not strictly depend on the special mechanism debate unless one takes a restrictive view of the special mechanism.

There is also a statistical difficulty because it is often not possible to distinguish between class and trait models. Bartholomew and colleagues (2002) point out that often the covariance matrix of observed variables can be equally well recreated by either assuming a set of discrete latent classes or by assuming continuous latent traits. This means that the models will fit equally well. When the latent classes are very different from each other (and within the classes they are relatively homogeneous), then it is possible to differentiate classes and traits. The taxometric approach (Ruscio, Haslam & Ruscio, 2006; Waller & Meehl, 1998; see also De Boeck, Wilson & Acton, 2005) has become a popular method for doing this in clinical psychology and it could be used also in FBM research. Within memory research a popular approach is to assume that there are two processes, one of which is continuous and one which is discrete. However, even with the large number of trials from the typical recognition memory study it is often difficult to distinguish this model from others (Malmberg, 2002) and much of the data are consistent with both processes being continuous (Wixted, 2007). Because of these difficulties FBM researchers should consider both the trait and the class models of FBMs.

How to measure FBMs?

Use multiple measures. Consider both trait and class models (latent variable models were discussed here, but the data reduction techniques discussed by Curci in Chapter 1 can also be used).

Summary on Measurement

Science requires several decisions about measurement. If FBM researchers want to advance beyond the trivial observation that some people have good memories (which have various correlated attributes) for some news events, it is necessary to be explicit about:

(1) What the population under investigation is,
(2) What distinguishes FBMs from other events, and
(3) How this variable should be measured?

Research needs to be conducted to describe the distribution which is naively shown in Figure 2.2 as a normal distribution. At present, very little can be said about this distribution, which makes any scientific statements about FBMs difficult to interpret.

TYPES OF FBM HYPOTHESES

There is some population, P, which the researchers want to make inference about. It is usually impossible to study this entire population, and therefore a sample, S, is taken from this. Making inference about a population from data of a sample has a long tradition (for example, on the first page of Gossett’s (1908, p. 398) classic t test paper: any result “is only of value in so far as it enables us to form a judgment as to the statistical constants of the population”), but requires certain rules. Depending on whether the researchers assume S is representative of P, there are two types of inference that can be made from data of S about P. If S is assumed to be representative of P, then providing sampling and measurement error are taken into account the researchers can make positive statements about all of P. For example, suppose you have a sample of memories that you think are representative of the population of interest, and find the correlation between the emotional reaction at the time of the event and the clarity of the subsequent memory has a 95% confidence interval of 0.3 to 0.4. This would allow you to conclude, with some confidence (although what exactly “confidence” means is another thorny issue), that the correlation in the population is within this band. Suppose instead that you are not willing to assume that the sample is representative of the population either because the people chosen are not representative of the population of people or the memories chosen are not representative of the population of these people's memories. Instead you assume the sample is just representative of some (often unknown) subset of the population. Finding an interval of 0.3 to 0.4 only allows you to say that you are confident that the correlation is in this range for this particular subset.

This is called local inference (Lunneborg, 2000). If someone has put forward the hypothesis that "for all subsets of the population the correlation is 0.5", then the finding that this does not hold for any particular subset allows this "for all subsets" hypothesis to be rejected. Thus, the hypothesis that the motions of all bodies conform to Newton's laws can be rejected by precise observation of the orbit of the planet Mercury. This is why Popperian falsification is used in many sciences. However, while "for all subsets" hypotheses are popular in some of the sciences, it is difficult to argue for their applicability with reference to FBMs. It is likely that people would expect correlations to vary in different subsets of a population (as is usually true in the medical and social sciences; see Engels, Schmidt, Terrin, Olkin & Lau, 2000).

Both of these are associative hypotheses, in that they are about the population joint distribution (often measured by the correlation) of emotion and clarity for either the entire population or for some subset of the population. They are not about causality, although as discussed below they can be used to inform causal theories. There are questions about the role of associative hypotheses in science (Cronbach, 1957; Fodor, 1991; Spearman, 1904; Wright, 2006), but it is clear that particularly within FBM research associate hypotheses are often discussed.

The sample, S, is often divided into groups. Suppose there are two groups, S1 and S2. This division could be based on something outside of the control of the researcher, like whether the participants were watching TV when some FBM event occurred or the participants' gender. In these cases it is a "quasi-experiment" and associative inferences are usually made (i.e., emotion and clarity are associated, rather than emotion causes clarity). In some cases researchers might use the associative finding (TV is associated with a certain types of memories) to infer some causal relationship (the graphic TV images cause certain types of memories), but this requires further assumptions (like those watching the television are similar to those not watching the television). In other cases making causal inference from an association is not possible (for example, gender cannot cause anything, see Holland, 1986).

If the division into S1 and S2 is random, as often occurs with experiments, the researchers are able to make causal inferences more easily, albeit with some caveats (Cook & Campbell, 1977). The critical aspect of drawing causal inferences is that the two groups should differ only by sampling error (which can be estimated if random assignment is used) and whatever the researchers want to make the causal inference about. The philosophers call this the ceteris paribus conditional, and as Cook and Campbell (1979, p. 5, emphasis in original) put it: "random assignment is the great ceteris paribus–that is, other things being equal–of causal inference." It is not possible to randomly allocate somebody into a condition where they think, for example, that the death of Princess Diana was an important event. Most FBM studies compare naturally occurring groups which may differ in many ways other than just what they thought about the event. In fact, it is the aim of much sociology and social psychology to document which groups of people are most affected by particular events. To make causal inference about, for example, thinking her death was an important event, requires some assumptions. While it would clearly be wrong for researchers to assume that, for example, people who thought Diana's death was a very important event were the same as others in all other ways, a researcher might feel that it is plausible to assume that any of the other differences are not associated with the dependent variable, memory quality. This assumption is of course testable if enough of the other possible confounding variables have been accurately measured[5].

The purist might argue that because FBM data often cannot unequivocally lead to causal conclusions, that these data are of no value in reaching causal conclusions. This narrow and short-sighted view would also send much neuropsychology and astronomy data to the dustbin. I am often asked by students if they are allowed to put forward some causal theory after collecting data from a quasi-experiment as if the causal theory police are on the prowl. People put forward causal theories with no data (or in the case of Brown and Kulik [1977], with some of their data counter to their theory, footnote 1) without fear of arrest, so having some data to help guide theory construction is an improvement!

One of most over-used mantras of first-year psychology statistics is "correlation does not imply causation". Strictly speaking, finding a correlation between, say, emotion and clarity, does not imply that emotion causes clarity, but finding a correlation does show that there almost certainly exists some causal relationships between various parts of the complex network of hypotheses within which these two attributes are embedded (see Meehl & Waller, 2002; Wright, 2006). If this network is viewed spatially, where the number of intervening nodes is a measure of distance, then because all other things being equal, simpler models should be preferred to more complex models (i.e., Occam's razor), it is best to consider causal relationships between the nearest events as more likely than more distant relationships. Further, because some events are separated in time (e.g., emotion at the time of an event and clarity of a subsequent memory), and outside of some physics hypotheses causes precede effects, this can also help in locating possible causal relationships.

FBM research is usually correlational/quasi-experimental. The data are about associations. While specific causal statements can be made on the basis of correlational data, there are often other plausible explanations including that some other variables cause variation in both attributes (Simon, 1954). For example, while most FBM/autobiographical memory research shows that emotion and clarity are positively correlated, it appears this may be due to the event's importance being causally related to both (Wright & Nunn, 2000). This is supported by numerous laboratory studies which indicate that heightened emotion negatively affects memory clarity (Deffenbacher, Bornstein, Penrod & McGorty, 2004). This is an example where the associative hypothesis (that emotion and clarity are positively related) appears in the opposite direction of the causal hypothesis (that emotion impairs memory).

While causal inference from correlational data is possible, it has numerous pitfalls, it is open to alternative interpretations, and requires caveats and assumptions. This does not mean that the data from a typical FBM study are not valuable, much of them are, but it does question whether alternative designs could address many of the causal questions in better ways.

The Case Study Approach

Any discussion of FBM methods requires some mention of the case study approach to science. Case studies are a popular method in neuropsychology where the case is the individual person, but the phrase "case study" applies also to where the case is an individual event (Wells & Windschitl, 1999). Case studies are an extreme example of where the sample is representative of only a subset of the population, but they can be useful. For example, a typical method in long-term memory research with children is to show them the same event and then later test them on this event, comparing either across experimental or non-experimental groups. By using the same event it provides some control for the study. However, it does mean that the researcher needs to be cautious about generalizing to other events. The difficulty with the single event approach with FBM research is that differences in the events are often the key variables under investigation. Therefore, controlling the event is counter-productive.

If a researcher is interested in making inferences about all events, studying only, for example, the Hillsborough Football disaster limits what can be said about all events. Wright (1993) found systematic differences in people's memories of this disaster across time. He could not conclude anything in particular about events, even those of FBM calibre, only that one particular event, which was arguably of FBM calibre for many of his participants, yielded memories incompatible with the hypothesis: "for all subsets of events of flashbulb calibre, there are no systematic biases in memories".

Wright (1993) does say something about the Hillsborough event. As such, it is of interest to people who actually want to know about the event. This is different from the aims of most cognitive psychologists, but in line with researchers from other disciplines. Some of the best case studies relevant to FBMs are by anthropologists studying memories for some of the bizarre and often horrific rituals in other cultures (Whitehouse, 2000). There have also been valuable studies of particular political and cultural events. In one of the best studies of FBMs, Judith Zur (1998) examined La Violencia, the violence surrounding the ethic war in Guatemala. The authorities forbid discussion of the atrocities: "The entire history of la violencia can be read as a war against memory, an Orwellian falsification of memory, a falsification of reality" (p. 159). Zur's research provides great insight into this often forgotten war, but it also provides insight to the workings of memory in a situation that it would be immoral to mimic in a laboratory. She provides graphic illustrations that memories, some accurate and some distorted, prevail even when not being allowed to talk about the kidnappings and murders. Zur's work provides an example, outside of cognitive psychology, of using the case study approach applied to memory.

As the case study approach is relatively rare within cognitive psychology (outside of neuropsychology where the individual is often the case), the question is why it became the method of choice for FBM researchers. Brown and Kulik's (1977) original study used multiple events and this allowed them to make valuable comparison across events. They used part of these data as evidence for their model. Although as discussed in footnote 2 some of their data do not support this model, their design was a good approach. Most cognitive psychologists are interested in how different events elicit different types of memories. Therefore, it is surprising that the FBM literature has tended to use single events, or a small number of events, rather than asking about several events which differ by the attributes of interest.

I believe there are two reasons that cognitive psychologists have opted for the case study approach when researching FBMs. The first is the desire to examine consistency (or accuracy). In order to do this, researchers have to question participants soon after an event, and this is only practical for a single event. The second reason is that since Neisser's (1982) and McCloskey, Wible and Cohen's (1988) case studies which effectively falsified the Brown and Kulik's (1977) bold "for all sets" conjecture that FBMs are very accurate, this method has gained acceptance beyond reproach as the way to study the phenomenon in general, rather than a method limited to falsifying bold conjectures. The next major news event will trigger a series of studies on FBMs. It seems that FBM studies often begin with a news event rather than a research question. The growing awareness of the FBM concept has made the study of FBMs acceptable in its own right. Figure 2.3 shows that the citations of Brown and Kulik and use of the phrase "flashbulb memory" have increased dramatically in recent years. The study of FBMs has become its own area of research, but as mentioned in the first half of this paper, it has become an area of research without it being clear what FBMs are. This is an unfortunate situation.

FIGURE 2.3. NEAR HERE

Summary on Hypotheses

There are two broad types of hypothesis: causal and associative. The design of the study should be congruent with the type of hypothesis of most importance because it is difficult to make inference of one type from a study designed to investigate the other type. Wright (2006) describes the differences between these types of hypotheses and their associated methods with reference to eyewitness testimony research, and many of the same issues apply. FBM researchers also have to decide whether they are interested in differences between people or between events. For example, Wright, Gaskell and O'Muircheartaigh (1998) were explicit that they were interested in group differences. Therefore, they asked a large representative sample of UK adults about a couple of events so that they could make positive statements about the associations for these events. If interest is in group differences, then this is the right approach. If the interest is in how somebody would remember different types of events, then the better approach is to ask a relatively small number of people about a large number of events (Wright & Gaskell, 1995), as is often done in other areas of autobiographical memory research (Wright & Nunn, 2000).

CONCLUSION

As a postgraduate at the London School of Economics (LSE), I was researching what had become known as everyday memories. I would go to library and dutifully photocopy lots of papers, some of which I would read. So that I could read without interruption, I would go off to other departments' seminar rooms. One day I found myself in the philosophy seminar room under a picture of Imre Lakatos reading Banaji and Crowder's (1989) attack on everyday memory research; I was angry. How dare somebody attack the sacrosanct topic on which I was conducting my PhD research!

The irony of my physical location only became apparent in subsequent reads. While at the LSE, Lakatos (1977) had argued that research programs were either degenerative or progressive. Each had a core of beliefs that remained unchallenged. In the case of the everyday memory movement, Banaji and Crowder were arguing that the everyday memory movement was degenerative: "no theories that have unprecedented explanatory power have been produced; no new principles of memory have been discovered, and no methods of data collection have been developed that add sophistication or precision" (p. 1185).

I now view the Banaji and Crowder (1989) paper as a very positive for the study of everyday memories. It forces people to step back and ask questions about the methods they use, and they pointed people towards ways of making everyday memory research generalizable. Just because people use a particular type of stimuli does not excuse them from asking if there are better methods. It is important to make sure that FBM research avoids being in a degenerative rut. As Figure 2.3 shows FBM research is an increasingly popular topic, but this would be problematic if its popularity protected researchers from exploring the core of what it is to be a FBM. It is necessary to develop a theory of measurement that allows FBMs to be studied in relation to other memories and to use methods that are appropriate to the research questions. The current research has shown that there exist some events that produce good memories and some events that produce bad memories. FBM researchers need to use other methods to go beyond this initial statement.

As said at the beginning of this paper, I am playing devil's advocate to a certain degree. This book is testament to the study of FBMs has made some progress. But more progress can be made.

NOTES

[1] This chapter was prepared while Dan Wright was on sabbatical at Florida International University, where he is now full-time. Much thanks to Antonietta Curci and several reviewers for useful comments on previous drafts.

[2] This statement deserves further explanation. Brown and Kulik's (1977) main group comparisons, which they use to justify their emotion combined with consequentiality produces FBMs model, were that Black people had more FBMs than White people for civil rights events, and these events were more consequential events for Black people than for White people. However, in their study Black people also rated both Robert Kennedy's assassination and Ted Kennedy's Chappaquiddick incident as more consequential than White people did, but Black people had fewer FBMs for these events. Similarly, Black people rated Martin Luther King's assassination as more consequential than John F. Kennedy's, but had fewer FBMs for MLK's assassination. See Wright, Gaskell and O'Muircheartaigh (1998) for further examples of this type. These observations do not falsify the Brown and Kulik model, but this difficulty in falsifying a causal model with correlation data is a theme of this chapter and one of the reasons why the amount written about FBMs exceeds the amount of scientific progress in understanding them.

[3] There are lots of other possibilities. For example, the population in the typical memory recognition study is all events that participants are given the opportunity to recognize.

[4] Beginning with Neisser (1982), there has been much discussion about how accurate responses are to these questions. This is why researchers now usually ask people at two or more points in time. This produces a measure of consistency, rather than accuracy. Winningham, Hyman and Dinnel (2000) discuss these issues and stress that if researchers want to use consistency as a proxy for accuracy it is important to gather initial reports very soon after the event (see also Coluccia, Bianco, & Brandimonte, 2006).

[5] Another design that is sometimes used compares some event which the researchers feel is of FBM calibre with some event which they do not feel is of this calibre. These events usually vary on many variables, including presumably whatever the x-axis of Figure 2.2 is supposed to be. If the researchers wish to claim that differences on any dependent measures are due to some flashbulb-ness of the "FBM event", then they either have to claim that the events only differ on this flashbulb-ness or that no other differences are likely to relate to the dependent measures. It is unlikely that either of these are valid assumptions, thus limiting the inference that can be made from these designs.

REFERENCES

Banaji, M.R. & Crowder, R.G. (1989). The bankruptcy of everyday memory. American Psychologist, 44, 1185-1193.

Bartholomew, D. J. & Knott, M. (1999). Latent variable models and factor analysis (2nd ed). London: Arnold.

Bartholomew, D. J., Steele. F., Moustaki, I., & Galbraith, J. I. (2002). The analysis and interpretation of multivariate data for social scientists. London: Chapman & Hall/CRC.

Brewin, C. R. (2003). Posttraumatic stress disorder: Malady or myth? New Haven, CN: Yale University Press.

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

Coluccia, E., Bianco, C., & Brandimonte, M. A. (2006). Dissociating veridicality, consistency, and confidence in autobiographical and event memories for the Columbia shuttle disaster. Memory, 14, 452-470.

Conway, M. A., Anderson, S. J., Larsen, S. F., Donnelly, C. M., McDaniel, M. A., McClelland, A. G. R., Rawles, R. F., & Logie, R. H. (1994). The formation of flashbulb memories. Memory & Cognition, 22, 326-343.

Cook, T.D. & Campbell, D.T. (1979). Quasi-experimentation: Design & analysis issues for field settings. Boston: Houghton Mifflin Company.

Cronbach, L.J. (1957). The two disciplines of scientific psychology. American Psychologist, 12, 671-684.

Curci, A. (2005). Latent variable models for the measurement of flashbulb memories: A comparative approach. Applied Cognitive Psychology, 19, 3-22.

De Boeck, P., Wilson, M., & Acton, G. S. (2005). A conceptual and psychometric framework for distinguishing categories and dimensions. Psychological Review, 112, 129-158.

Deffenbacher, K. A., Bornstein, B. H., Penrod, S D., & McGorty, E. K. (2004). A meta-analytic review of the effects of high stress on eyewitness memory. Law and Human Behavior, 28, 687-706.

Engels, E.A., Schmidt, C.H., Terrin, N., Olkin, I., & Lau, J. (2000). Heterogeneity and statistical significance in meta-analysis: An empirical study of 125 meta-analyses. Statistics in Medicine, 19, 1707-1728.

Festinger, L., Riecken, H. W., & Schacter, S. (1956). When prophecy fails: A social and psychological study of a modern group that predicted the destruction of the world. Originally published Minneapolis: University of Minnesota Press (Harper and Row). Now available from APA print-on-demand at http://books.apa.org/.

Finkenauer, C., Luminet, O., Gisle, L., El-Ahmadi, A., van der Linden, M., & Philippot, P. (1998). Flashbulb memories and the underlying mechanisms of their formation: Toward an emotional-integrative model. Memory & Cognition, 26, 516-531.

Fodor, J.A. (1991). You can fool some of the people all of the time, everything else being equal; Hedged laws and psychological explanations. Mind, 100, 19-34.

Gardiner, J. M. (1988). Functional aspects of recollective experience. Memory & Cognition, 16, 309-313.

Gardiner, J. M., Ramponi, C., & Richardson-Klavehn, A. (2002). Recognition memory and decision processes: A meta-analysis of remember, know, and guess responses. Memory, 10, 83-98.

Gossett, W., writing as “Student” (1931). The Lanarkshire milk experiment. Biometrika, 23, 398-406.
Hand, D.J. (2004). Measurement theory and practice: The world through quantification. London: Edward Arnold.
Holland, P.W. (1986). Statistics and causal inference. Journal of the American Statistical Association, 81, 945-960.
Lakatos. I. (1977). The methodology of scientific research programmes: Philosophical papers (Vol. 1). Cambridge, UK: Cambridge University Press.
Livingston, R. B. (1967). Reinforcement. In G. C. Quarton, T. Melnechuck, & F. O. Schmitt (Eds.), The neurosciences: A study program (pp. 568-576). New York: Rockefeller University Press.
Lunneborg, C. E. (2000). Data analysis by resampling: Concepts and applications. Pacific Grove, CA: Duxbury Press.
Malmberg, K. J. (2002). On the form of ROCs constructed from confidence ratings. Journal of Experimental Psychology: Learning, Memory, and Cognition, 28, 380-387.
McCloskey, M., Wible, C. G., & Cohen, N. J. (1988). Is there a special flashbulb-memory mechanism. Journal of Experimental Psychology: General, 117, 171-181.
McNally, R. J. (2003). Remembering trauma. Cambridge, MA: Harvard University Press.
Meehl, P.E. & Waller, N.G. (2002). The path analysis controversy: A new statistical approach to strong appraisal of verisimilitude. Psychological Methods, 7, 283-300.
Moustaki, I. (1996). A latent trait and a latent class model for mixed observed variables. British Journal of Mathematical and Statistical Psychology, 49, 313-334.
Neisser, U. (1982). Snapshots or benchmarks? In Neisser, U. (Ed.), Memory observed: Remembering in natural contexts (pp. 43 48). San Francisco: Freeman.
Rubin, D. C. (1992). Definitions of autobiographical memory. In M. A. Conway, D. C. Rubin, H. Spinnler, & W. A. Wagenaar (Eds.), Theoretical perspectives on autobiographical memory (pp. 495-499). The Netherlands: Kluwer Academic Publishers.
Ruscio, J., Haslam, N., & Ruscio, A. M. (2006). Introduction to the taxometric method: A practical guide. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
Simon, H.A. (1954). Spurious correlation: A causal interpretation. Journal of the American Statistical Association, 49, 467-479.
Spearman, C. (1904). “General Intelligence”: Objectively determined and measured. American Journal of Psychology, 15, 201-93.
Tulving, E. (1985). How many memory systems are there? American Psychologist, 40, 385-398.
Waller, N. G., & Meehl, P. E. (1998). Multivariate taxometric procedures: Distinguishing types from continua. Thousand Oaks, CA: Sage.
Wells, G.L., & Windschitl, P.D. (1999). Stimulus sampling in social psychological experimentation. Personality and Social Psychology Bulletin, 25, 1115-1125.
Whitehouse, H. (2000). Arguments and icons: Divergent modes of religiosity. Oxford: Oxford University Press.
Winningham, R. G., Hyman, I. E., & Dinnel, D. L. (2000). Flashbulb memories? The effects of when the initial memory report was obtained. Memory, 8, 209-216.

Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. Psychological Review, 114, 152-176.
Wright, D. B. (1993). Recall of the Hillsborough disaster over time: Systematic biases of 'flashbulb' memories. Applied Cognitive Psychology, 7, 129-138.
Wright, D. B. (2006). Causal and associative hypotheses in psychology: Examples from eyewitness testimony research. Psychology, Public Policy, and Law, 12, 190-213.
Wright, D. B., & Gaskell, G. D. (1995). Flashbulb memories: Conceptual and methodological issues. Memory, 3, 67-80.
Wright, D. B., Gaskell, G. D., & O'Muircheartaigh, C. A. (1998). Flashbulb memory assumptions: Using national surveys to explore cognitive phenomena. British Journal of Psychology, 89, 103-122.
Wright, D. B., & Nunn, J. A. (2000). Similarities within event clusters in autobiographical memory. Applied Cognitive Psychology, 14, 479-489.
Zur, J. N. (1998). Violent memories: Mayan war widows in Guatemala. Oxford: Westview Press.

FIGURE CAPTIONS

Figure 2.1. A simplified FBM model, based on Brown and Kulik (1977) and subsequent models.

Figure 2.2. A distribution, for some population of events, of memorability with events beyond some value being called FBMs.

Figure 2.3. The number of times Brown and Kulik (1977) was cited each year since publication (shown with a +) and the number of times the phrase “flashbulb memory” or “flashbulb memories” was used in titles and abstracts (shown with a ●). The data are from the ISI Web of Knowledge collected on 13 March, 2008 and go up to 2007.

Figure 2.1.

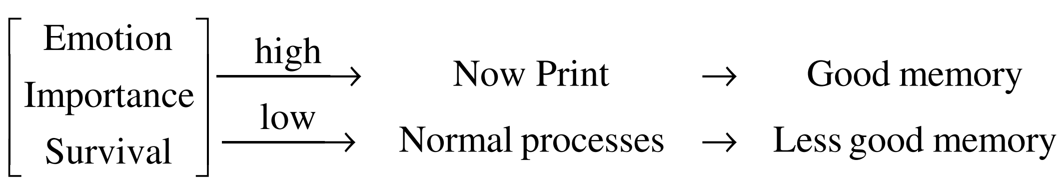

Figure 2.2.

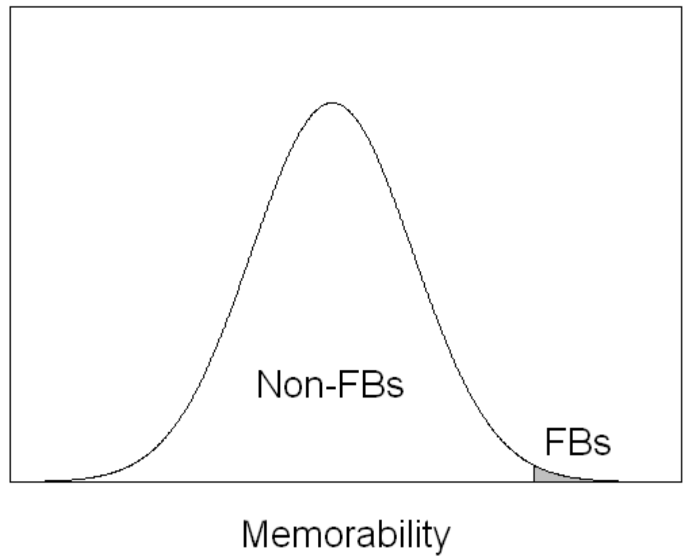

Figure 2.3.