Cognition, 46 (1993) 101–128

Reconstructive remembering of the scientific literature*

Kim J. Vicente
Department of Industrial Engineering, University of Toronto, Toronto, Ontario M5S 1A4, Canada

William F. Brewer
Psychology Department, University of Illinois at Urbana-Champaign, Champaign, IL 61820, USA

Received March 23, 1992, final version accepted August 27, 1992

Abstract

Vicente, K.J., and Brewer, W.F., 1993. Reconstructive remembering of the scientific literature. Cognition, 46: 101–128.

In this paper we investigate the role of reconstructive memory in citation errors that occur in the scientific literature. We focus on the case of de Groot's (1946) studies of the memory for chess positions by chess experts. Previous work has shown that this research is very often cited incorrectly. In Experiment 1 we show that free recall of this work by research psychologists replicates most of the errors found in the published literature. Experiment 2 shows that undergraduates reading a correct account of the de Groot study also make the same set of errors in recall. We interpret these findings as showing that consistent errors in secondary accounts of experimental findings are frequently reconstructive memory errors due to source confusion and schema-based processes. Analysis of a number of other examples of

Correspondence to: Dr. Kim. J. Vicente, Department of Industrial Engineering, University of Toronto, Toronto, Ontario M5S 1A4, Canada. (email: benfica@ie.utoronto.ca).

*We would like to thank the following people for their help in formulating the ideas presented here: Reid Hastie, Dedre Gentner, Sally Cohen, John Flach, Alex Kirlik, John Lee, Randy Sollenberger, Herbert Simon, and the members of the Cognitive Psychology Seminar Group at the University of Illinois. We also want to thank Reid Hastie for leading us to several of the cases of mis-citation that occurred in the social psychology literature. A special thank you to Adriaan de Groot for invaluable comments and historical details, to Susan Perkins for help with data collection, to Ellen Brewer and Ann Bisantz for comments on an earlier draft, and to Penny Sanderson for her contribution to Experiment 1 and for her continued support and stimulation of the ideas presented in this paper.

scientific literature that have been frequently cited incorrectly add additional support to the reconstructive account. We conclude that scientists should be aware of the tendency of reconstructive memory errors to cause violations of the scientific norm of accurate reporting of the scientific literature.

RECONSTRUCTIVE REMEMBERING OF THE SCIENTIFIC LITERATURE

Most working scientists are aware that secondary sources frequently give accounts of experiments and results that are not accurate. For example, in a discussion of the influence of language on thought, Whorf (1940/1956) stated that the Eskimo language had a number of different words for snow (in a figure in the article, he gave the number as three and in the text he implies the number might be seven). However, many secondary sources discussing Whorf's work give a much larger number, with published accounts reaching as high as 100 different words for snow (Martin, 1986; Pullum, 1989).

Published examples of distortions in the scientific literature occur across disciplines, from the "hard" sciences to the "soft" sciences. Examples include: Evariste Galois' supposed invention of the branch of mathematics known as group theory on the night preceding the duel that lead to his death at the age of 20 (Rothman, 1989, pp. 148–193), the extent of the influence of the Michelson–Morley experiment on Einstein's development of special relativity (Holton, 1988, pp. 279–370), the origins of social psychology (Haines & Vaughan, 1979), the details of the psychoanalytic case history of Breuer and Freud's patient Anna O (Ellenberger, 1972), Galileo's purported experiment with falling objects from the Tower of Pisa (Cooper, 1935), and the way in which the Hawthorne experiments on worker productivity were conducted (Gillespie, 1988).

While there are clearly many different reasons for distortions in the scientific literature, we propose that one major source is the operation of schema-based reconstructive memory processes (cf. Bartlett, 1932; Brewer & Nakamura, 1984) in scientists. To support our general argument, we report an intensive analysis of one particular case of distortion in the scientific literature: the work of de Groot on memory in chess masters (cf. Vicente & de Groot, 1990).

First we describe the original de Groot publications and give an account of some of the distortions of this study that have appeared in the scientific literature. Then we report the results of a memory experiment with research psychologists confirming that some of these distortions are, in fact, widespread. We then propose a psychological explanation for these distortions, based on a theory of reconstructive remembering. Next, three experiments designed to test our memory theory are described. We then review other published cases of distortions to show that the reconstructive remembering explanation gives a good account of

these errors as well. Finally, we outline some general implications of reconstructive remembering for the conduct of scientific research.

DE GROOT'S STUDY OF MEMORY IN CHESS MASTERS

The particular scientific literature we have focused on originated with the seminal work of de Groot (1946, 1965) on problem solving in chess. If the reader is familiar with the de Groot research we recommend that the reader stop at this point and write out a brief summary of what they can recall of this work before reading the remainder of this paragraph. The well-known result to emerge from these studies is that, after being exposed to a chess position for a very brief time, chess masters are able to reconstruct the position with high accuracy, whereas the performance of lesser players is nowhere near as impressive. In subsequent work, it was demonstrated that when masters are compared with weaker players on memory for a chess board on which the pieces are arranged randomly, there is little difference between the two types of players. The finding from the control condition indicates that the masters' superior recall ability is not due to a general advantage in visual memory but specifically tied to their chess expertise (see Vicente, 1988, for a more comprehensive review of the literature on expertise effects in memory recall).

The work of de Groot (1946, 1965) has had a wide scientific impact. The *Social Sciences Citation Index* reveals that de Groot's monograph has received an average of 10.5 citations per year between 1966 and 1988. According to data cited by Price (1986), this places de Groot's work in the 99.9th percentile in terms of frequency of citation. Thus, it is evident that de Groot's research has had an unusually great influence on the literature.

TRACING THE ERRORS IN THE LITERATURE

Vicente and de Groot (1990) have pointed out that there has been a persistent distortion of some of the details of de Groot's (1946, 1965) original work in this area. To provide a context for the ensuing discussion of these errors, the historical details of the original studies of the memory of chess experts are briefly summarized below (see Vicente & de Groot, 1990, for a more detailed account).

A historical account of studies of chess memory

As part of his dissertation, de Groot (1946, 1965, pp. 321–334) conducted an experiment to test chess players' abilities to reconstruct a board position. There

were four subjects with different levels of chess expertise (grandmaster, master (de Groot himself), expert, and class level). Subjects were presented with mid- and end-game positions from actual games. The exposure times ranged from 2 to 15 s. De Groot found that the two more experienced players were able to score over 90% (as determined by an arbitrary scoring criterion). The recall performance of the less skilled subjects was considerably worse. Qualitatively similar results had been reported by Djakow, Petrowski, and Rudik (1927), although there were several methodological flaws in their experiment (cf. de Groot, 1965, p. 322). Many years later, de Groot's result was replicated by Jongman (1968), a student of his, this time with exposure times of 5 s. Some of Jongman's results appeared in a book chapter written by de Groot (1966). Jongman and a co-worker, Lemmens, also conducted a control experiment where the pieces were placed randomly on the board. The memory advantage found on authentic positions disappeared in this random condition. Unfortunately, this study was never published. In 1973, Chase and Simon (1973a, 1973b) replicated the original results of de Groot (1946, 1965) with novice and master level subjects and exposure times of 5 s. They also replicated the unpublished random condition findings of Jongman and Lemmens. This was the first published memory recall study that incorporated a random control condition.

Distortions in the published literature

Given this historical background, let us now look at the descriptions that Chase and Simon provided of de Groot's (1946, 1965) original memory recall experiment (see Table 1). While the Simon and Chase (1973) excerpt is, for the most part, historically accurate,[1] the reports of Chase and Simon (1973a, 1973b) are not. The passage from Chase and Simon (1973a), for instance, gives the impression that de Groot's (1946, 1965) experiment included a random control condition when in fact it did not. The passage from Chase and Simon (1973b), published in *Cognitive Psychology*, also gives the same erroneous impression.[2]

According to the *Social Sciences Citation Index*, Chase and Simon (1973b) is by far the most widely cited of the three papers, with an average of 16.4 citations per year. The next most cited paper is the *Visual Information Processing* chapter

[1]There is one oversight in this passage. The exposure times are reported as ranging from 2 to 10 s whereas the actual times adopted by de Groot (1946, 1965) ranged from 2 to 15 s. The reason for this discrepancy can probably be attributed to the wording and spatial layout of the printed text in de Groot (1965), which tends to encourage such a mistake: "Each position [//] had a prescribed exposure time, varying from two to ten seconds and [/] in one case as high as fifteen" (pp. 322–323). The symbols // and / indicate the beginning of a new page and new line, respectively.

[2]The exposure time of 5 s reported in this passage is not consistent with the 2–15 s reported in de Groot (1946, 1965), but it *is* in agreement with the exposure times reported in the other two publications cited in this passage (i.e., de Groot, 1966, and Jongman, 1968).

Table 1. *Three descriptions of de Groot's (1946, 1965) work*

Chase and Simon (1973a, pp. 215–217)
The early work on chess skill was done over 30 years ago by de Groot (cf. de Groot, 1965). He asked what it is that differentiates Master chess players from weaker players, and he studied some of the best chess players in the world at that time . . . [De Groot found] an intriguing difference between Masters and weaker players in their ability to perform a task involving perceptual and short-term memory processes. Masters were able to reconstruct a chess position almost perfectly after viewing it for only 5 seconds or so. There was a sharp drop-off in this ability for players below the Master level. This result could not be attributed to a generally superior visual short-term memory capability of the Masters because, when the pieces were placed randomly on the board, recall was equally poor for Masters and weaker players. Masters are subject to the same limitations on short-term memory as everyone else.

Chase and Simon (1973b, pp. 55–56): Historically inaccurate case used in Experiment 2
The most extensive work to date on perception in chess is that done by de Groot and his colleagues (de Groot, 1965, 1966; Jongman, 1968) . . . [De Groot found] an intriguing difference between masters and weaker players in his short-term memory experiments. Masters showed a remarkable ability to reconstruct a chess position almost perfectly after viewing it for only 5 sec. There was a sharp drop-off in this ability for players below the Master level. This result could not be attributed to the masters' generally superior memory ability, for when chess positions were constructed by placing the same numbers of pieces randomly on the board, the masters could then do no better in reconstructing them than weaker players. Hence, the masters appear to be constrained by the same severe short-term memory limits as everyone else (Miller, 1956), and their superior performance with "meaningful" positions must lie in their ability to perceive structure in such positions . . .

Simon and Chase (1973, p. 395): Historically accurate case used in Experiments 2 and 3
In Amsterdam, Adriaan de Groot . . . was the first psychologist to carry out extensive experiments on problem solving using chess as the task . . . He displayed a chess position to his subjects for a very brief period of time (2 to 10 seconds) and then asked them to reconstruct the position from memory. These positions were from actual master games, but games unknown to his subjects. The results were dramatic. Grandmasters and masters were able to reproduce, with almost perfect accuracy (about 93% correct), positions containing about 25 pieces. There was quite a sharp drop-off in performance somewhere near the boundary between players classified as masters, who did nearly as well as grandmasters, and players classified as experts, who did significantly worse (about 72%). Good amateurs (Class A players in the American rating scheme) could replace only about half the pieces in the same positions, and novice players (from our own experiments) could recall only about eight pieces (about 33%). There is a quite nice gradation on this perceptual task as a function of chess skill, and we have verified this in our own experiments (Chase and Simon, 1973).

We went one step further: we took the same pieces that were used in the previous experiment, but now constructed random positions with them. Under the same conditions, all players, from master to novice, recalled only about three or four pieces on average – performing significantly more poorly here than the novice did on the real positions. (The same result was obtained by W. Lemmens and R.W. Jongman in the Amsterdam laboratory, but their data have never been published.)

In sum, these experiments show that chess skill . . . can be detected easily using a perceptual task with meaningful chess content. The experiment with random boards shows that the masters' superior performance in the meaningful task cannot be explained in terms of any general superiority in visual imagery. The perceptual skill is chess specific.

(Chase & Simon, 1973a), with an average of 3.1 citations per year, followed by the *American Scientist* article (Simon & Chase, 1973), with only 2.1 citations per year. One therefore finds that while the least often cited of Chase and Simon's three papers (Simon & Chase, 1973) presents a historically accurate account, the most frequently cited paper (Chase & Simon, 1973b) incorrectly attributes the random control condition to de Groot. Although it is an innocent oversight, Chase and Simon's mis-citation has had significant ramifications.

Before these ramifications are pointed out, however, it is important to distinguish between four different classes of citation errors. *Attribution errors* are those where an experiment is attributed incorrectly. *Methods errors* are those where the experimental method is cited incorrectly. *Results errors* are those in which the experimental results are cited incorrectly. Finally, *gist errors* are those where the gist of the findings is cited incorrectly. Gist errors are, of course, the most serious. Chase and Simon's slips clearly fall into the first two categories: attribution (random control) and method (exposure times). However, many additional mis-citations of de Groot (1946, 1965) have since appeared in the literature. The following is a partial inventory of some of these errors, classified according to type.

With regard to errors of attribution, Chase and Simon (1973a, p. 217, 1973b, p. 56), Engle and Bukstel (1978, pp. 673–674), Gilhooly (1988, p. 61), Hayes (1978, pp. 191–192), Hofstadter (1979, p. 286), Hirsch (1988, pp. 61–62), Johnson-Laird and Wason (1977, p. 528), Moates and Schumacher (1980, p. 58), Saariluoma (1985, p. 385), Sanford (1985, p. 153), Schacter (1986, p. 195), Seamon (1980, p. 76), and VanLehn (1989, p. 565) all incorrectly attribute the origin of the random control condition to de Groot (1946, 1965). Summarizing, there is a total of at least 14 instances of attribution errors: three in review papers, two in popular works (both national best sellers written by academics), five introductory psychology texts, and four experimental papers.

Methods errors can also be found in the literature. For example, Schacter (1986, p. 195) states that de Groot (1965) presented subjects with a configuration of 32 pieces on the board (the maximum number of pieces in a chess game), when in fact the actual average number of pieces was considerably less since the positions were taken from mid- and end-games. Furthermore, there are also citations which misrepresent the exposure times used by de Groot (1946, 1965). For instance, Chase and Simon (1973b, p. 55), Gilhooly (1988, p. 61), Hofstadter (1979, p. 286), Posner (1988, p. xxx), Sanford (1985, p. 153), Seamon (1980, p. 76), VanLehn (1989, p. 565), and Vicente [sic!] (1988, p. 255) report exposure times of 5 s and Hirsch (1988, p. 61) of 5–10 s. The actual exposure times ranged from 2 to 15 s. Misrepresentations of the subjects' expertise are also prevalent. Several authors state that novice subjects were used (Hirsch, 1988, pp. 61–62; Hofstadter, 1979, p. 286; Posner, 1988, p. xxxi; Sanford, 1985, p. 153; Schacter, 1986, p. 195; Seamon, 1980, p. 76; VanLehn, 1989, p. 565). However, as

mentioned earlier, the recall studies conducted by de Groot (1946, 1965) included only four subjects, none of whom were novices. Altogether, there are at least 16 instances of methods errors distributed over nine publications. Four of these are review papers, two are popular works, and three are introductory psychology texts. No methods errors have been found in experimental papers.

Finally, there is also at least one instance of a results error (Hirsch, 1988, pp. 61–62) stating that the recall performance of the grand master was perfect, when in fact it was close to, but not quite at that level. This mis-citation occurred in a national best seller.

It is important to note that no gist error has been found in this literature. Despite the numerous mistakes listed above, all of the authors citing de Groot's work have conserved the gist of the overall findings, that is, that more experienced players recall more pieces than less skilled players on authentic positions, but that the skill advantage disappears when scrambled positions are presented. This result parallels the findings from Neisser's (1981) case study of John Dean's memory. The analysis also reveals that the severity of the errors has increased with time. Chase and Simon (1973a, 1973b) only committed an attribution error and a methods error (exposure times), but other methods errors (novice subjects, and number of pieces), and even a results error (perfect memory for masters) have since appeared in the literature. Finally, it appears that most of the mis-citations have appeared in sources such as review papers, introductory psychology texts, and popular works, rather than experimental papers.

DISTORTIONS IN THE PUBLISHED TEXTS

In this section we will take a closer look at the mis-citations just described, with the intent of developing a systematic explanation for these errors.

Distortions in Chase and Simon

It is clear that there are distortions in the citations of de Groot by Chase and Simon (1973a, 1973b) and that many authors obtained their knowledge of the experiment from these sources. However, this factor alone cannot explain all of the findings. First of all, several of the errors found in subsequent citations go beyond the original misrepresentation found in Chase and Simon (1973a, 1973b). Second, in at least one case (Gilhooly, 1988), it is clear that the author did read at least part of de Groot (1965) since data are cited that cannot be found in any other reference. Nevertheless, two citation errors were committed. While failure to read the original source may have caused some errors, it cannot account for all of the evidence.

The Matthew effect

Another factor which may have contributed to the mis-citations is the Matthew effect in science (Merton, 1968, 1988), which refers to the observation that "eminent scientists get disproportionately great credit for their contributions to science while relatively unknown ones tend to get disproportionately little for their occasionally comparable contributions" (Merton, 1988, p. 607). In this particular case, de Groot, who was a chaired professor at the University of Amsterdam, received credit for an experiment conducted by his more junior collaborators, Jongman and Lemmens. It is likely that the Matthew effect may have led to the original mis-citation by Chase and Simon, since Simon was aware of the results of both de Groot's and Jongman and Lemmens' studies (Vicente & de Groot, 1990).

Reconstructive remembering

The mis-citations in the literature can also be interpreted as an exemplary case of reconstructive remembering similar to the errors observed by Bartlett (1932) in his classic study of the reconstructive memory process. In one of the experiments reported by Bartlett, a story was given to a subject to remember, and then that subject's recollection was passed along in writing to someone else to remember, and so on. The story became increasingly distorted as it was disseminated to more and more people. The key point, however, is that the distortions were not random or chaotic, but instead occurred in a regular pattern. The following features in subjects' recall were observed: condensation, elaboration, invention, simplification, integration towards greater coherence, and omission of qualifications. Globally, these patterns tended to make the original story simpler and more meaningful, thereby making it easier to remember. Similar findings were also obtained by Allport and Postman (1945) in their investigation of the psychology of rumor.

According to Bartlett, these changes in the text are generated by a process of reconstructive remembering. That is, people retain a general impression of the original event and then use that impression to generate a simple, yet plausible, account of the forgotten details. Stated more eloquently: "So long as the details which can be built up around [the remaining impression] are such that they would give it a 'reasonable' setting, most of us are fairly content, and are apt to think that what we build we have literally retained" (Bartlett, 1932, p. 176). The reconstructive nature of everyday memory has since been well documented (see, for example, the various studies in Neisser, 1982, and Neisser & Winograd, 1988).

While each of the factors just described may be responsible for some of the distortions observed in the literature, the remainder of this paper will focus on the reconstructive remembering hypothesis.

EXPERIMENT 1

Our first memory experiment was designed to discover the degree to which the details of de Groot's experiment were distorted in the memories of a group of research scientists. Do we observe the same errors found in the literature?

Method

Psychology faculty members at two major research universities were given an open-ended question asking them to describe what they remembered about the details of de Groot's work, without consulting any references. A total of 15 faculty members responded.

Results

The recalls were scored very liberally; for example, if subjects reported the exposure times as a few seconds (instead of the actual 2–15 s), or if they reported the subjects' levels of expertise as expert and less skilled (instead of the actual chess rankings), their answers were classified as being correct.

The results are presented in Table 2. There were two interesting systematic recall errors. Many of the research psychologists made errors in describing de Groot's subjects' levels of expertise. All of the errors were of the same form; they recalled that the control subjects were novices. This mistake was also frequently observed in the review of the errors in the published literature. Note that this error was *not* present in any of the Chase and Simon papers. The other major systematic error was in recalling that de Groot used a random control condition. This error was present in the two most widely cited Chase and Simon papers but not in their third paper (Simon & Chase, 1973).

Table 2. *Experiment 1: Recall of de Groot (1946, 1965) by a sample of research psychologists*

	Percentages[a]		
Type of response	Correct	Error	Omit and other
Random control	13	80	7
Exposure times	20	13	67
Subjects' expertise	47	53	0

[a] $n = 15$.

Discussion

This experiment shows very clearly that the memories of a sample of research scientists for the de Groot study contain a number of systematic distortions. However, the design of this experiment does not allow us to isolate the locus of the distortion. The data are consistent with three different positions: (a) the sources these scientists read were distorted, but the memories of the scientists were accurate representations of these texts; (b) these scientists had read one of the correct sources and this information became distorted in the recall process; and (c) there were distortions in both the original texts and in the memory of the scientists. The second author of this paper was one of the subjects in this sample and made memory distortions. He had read de Groot (1965), so we know that for at least one scientist in the sample we can exclude option (a). In Experiment 2 we attempt to distinguish these alternatives. However, first, we will give a more detailed account of the processes at work in reconstructive memory.

RECONSTRUCTIVE MEMORY IN SCIENCE

Reconstructive memory is the result of a complex set of psychological processes. However, for the purposes of the present paper we will focus on two sources of reconstructive errors: (a) schema-based processes and (b) interference- and source-confusion processes.

Schema-based processes

Bartlett (1932) postulated that there are global knowledge structures (schemata) which interact with incoming instances to produce systematic distortions in recall (cf. Brewer & Nakamura, 1984). In the case of recalling scientific experiments, we hypothesize that scientists have a schema for the canonical scientific experiment which incorporates the belief that the ideal scientific experiment provides a clean test of a particular theory with appropriate controls and clear, unambiguous results. In this schema, the theory explains the findings, the controls reduce alternate interpretations, and the results provide clear-cut evidence in support of the theory. We further postulate that when scientists are attempting to describe a particular experiment to make a certain point or as an example for teaching purposes (e.g., in lectures or introductory texts), there will be a strong tendency for the idealized experiment schema to override the actual details of a given experiment.

Can this hypothesis account for the distortions of the de Groot study? If one compares the recurring errors in some of the published accounts with the actual historical details, the consistency with this schema-based explanation is striking.

For example, consider the description given in VanLehn (1989):

> de Groot (1965) [demonstrated] that chess masters could recall almost all the pieces and positions of a chess board after having seen the board for only five seconds. Novices could only recall a few pieces. Moreover, if the stimulus was a chess board with the pieces arranged randomly, the recall of the expert sank to the level of a novice. (p. 565)

First, and most obvious, the distorted version is a concise description of the original. Several details have been omitted. Second, the gist of the findings has been conserved. Third, there is also a greater degree of coherence. Rather than itemizing the details of exactly who did what over the years, all of the findings (including those of the random control) are neatly attributed to a single person, de Groot. This provides a cleaner, easier to remember account of the basic findings than the historically correct version. The types of errors in this distorted citation are remarkably similar to the pattern of schema distortions observed by Bartlett.

The postulated schema for canonical experiments provides a good account of the distortions in the published literature. First, one would predict that the scientists would remember the theoretical "point" of the experiment (i.e., that chess experts remember more than weaker players when the stimulus is meaningful, but not when the stimulus is random), which they do. Second, one would also predict that the random control experiment would be recalled along with the experiment with meaningful positions because the control condition is essential to the interpretation of the findings. After all, if the random control is not included, then it could be the case that experts simply have a better overall memory than lesser players. This aspect of the experiment schema would then account for the widespread attribution of the control condition to de Groot. Third, one would also predict that scientists would often recall that the control subjects were novices, since the vast majority of expertise studies involve a comparison between extreme groups. Thus, the novice subject group provides a canonical control group for studies of expertise.

The only frequently observed distortion which does not seem to be caused by reconstruction from an idealized experiment schema is the error in exposure times (i.e., 5 s instead of the actual 2–15 s adopted by de Groot). Thus, this error is probably caused by other types of psychological processes.

This analysis has demonstrated that the schema-based reconstruction hypothesis provides a plausible account of many of the distortions that were observed in the literature and in our sample of research psychologists.

Interference and source-confusion processes

It is known that in recalling information people sometimes forget the source of the information or attribute the information to the wrong source (cf. Schacter,

Harbluk, & McLachlan, 1984). This also fits our data very well. Some of the research psychologists in our sample may well have read about the use of a random control group in one of the Chase and Simon papers, forgotten the source of this information, and attributed it to the de Groot experiment.

Another memory phenomenon that is relevant to this work is the research showing the impact of a new piece of information on similar information already in memory (e.g., seeing a stop sign in a picture and later being told that the sign was a yield sign). There is a large body of evidence showing that misleading postevent information can strongly affect subjects' later memory performance (e.g., Loftus, Miller, & Burns, 1978; McCloskey & Zaragoza, 1985). Note that much of this literature is focused on the issue of the form of the underlying memory *representation*; however, for our purposes, we are simply interested in the evidence that these processes cause distortions of memory *performance*. Once again this memory phenomenon is consistent with our data. It is likely that, at one time, some of our research psychologists read de Groot's monograph and then later read Chase and Simon's distorted account of de Groot (1946, 1965), and that the latter reading led to interference when reconstructing the details of de Groot's study.

We think that all of these processes contributed to the errors seen in reconstructive recall. Sometimes it seems possible to conclude that one of the processes is playing the major role in causing the errors. For example, the distortions of the number of words for snow in the Eskimo language (discussed earlier) seem driven by the ideal science schema process (e.g., the more words for snow there are the stronger is the evidence for the Whorf–Sapir hypothesis). On the other hand, the particular circumstances of the discussions of the de Groot experiment in the scientific literature make it a potential candidate for source confusions and interference effects. Note, however, that in this case the consistent attribution of the random control to de Groot (and never the attribution of the de Groot methodology to Case and Simon) also suggests the influence of the idealized science schema.

In the next section, we describe an experiment conducted to study empirically our hypothesis about the reconstructive recall of science texts.

EXPERIMENT 2

If it is the case that reconstructive remembering is responsible for the errors observed in the citations and in our sample of scientists, then it should be possible to replicate these same errors in a memory experiment. In this experiment, undergraduate subjects were asked to read one of two texts, and after a short delay were asked to recall the material they had read. The two texts were taken from two papers by Chase and Simon. One text provides a historically inaccurate description of de Groot's experiment, while the other text provides a historically

correct description of that study. If most of the distortions occur in the writers of secondary accounts and not in the recall of these accounts, then one would expect that errors would be found *only* in subjects reproducing the passage that provides an erroneous description of de Groot's work. However, if reconstructive recall occurs for both writers and readers, then one would expect that some of the subjects who are presented with the historically accurate text will commit the same reconstructive errors observed in the literature. It is important to note that undergraduate subjects may have an undeveloped experiment schema and so could show less impressive schema-based memory errors than scientists.

Method

Subjects

There were 34 subjects, who were all undergraduates at the University of Illinois and participated in the experiment as part of the requirements for introductory psychology.

Materials

The two texts used in this experiment were taken from the published literature. One provides a distorted account of de Groot's experiment (Chase & Simon, 1973b) and the other provides a historically accurate account of de Groot's experiment (Simon & Chase, 1973). The two texts are given in Table 1.

Design

There were 17 subjects in each condition. Each subject was presented with one of the two passages.

Procedure

At the beginning of the experiment, subjects in both conditions were provided with the following set of instructions:

> This is an experiment concerned with assessing your ability to evaluate scientific research. You will be given a short period of time to read a passage describing an experiment and then, when the experimenter tells you, you are to turn the page and answer several questions about the passage you just read. You are to answer the questions without looking back at the passage so please use your time to study the passage carefully. After the first passage you will receive several more passages with the same procedure.

Subjects were then given 3 min to read over the passage. The page with the text had the following explanatory statement at the top:

> This passage is about psychological work conducted on skill in chess. The passage was taken from a scientific paper written in 1973 by two authors, Chase and Simon, who were summarizing the findings from the work that had been done in this area up to that point.

After the 3 min were up, subjects were given 1 min to answer a few simple questions about the passage they had just read. These questions were designed to be neutral with respect to the memory errors and were included to make sure the subjects were motivated to comprehend the text. The questions were: (1) "The chess positions were from games unknown to the subjects." [T] [F]. (2) "The masters' superior performance is due to superior visual memory." [T] [F] . This was followed by approximately 30 min of unrelated activity.

After this time interval, subjects were asked to write down as much as they could about the passage they had read earlier. It should be noted that the passage from Simon and Chase (1973) contained a description of both de Groot's and Chase and Simon's work, whereas the passage from Chase and Simon (1973b) did not describe Chase and Simon's own work. Because the purpose of the experiment was to determine whether subjects would misrepresent the details of de Groot's experiment, those subjects receiving the passage from Simon and Chase (1973) were given very explicit instructions so that there would not be any confusion in the minds of these subjects about what information they were to recall. These instructions stated:

> We do not want you to recall the whole passage. We only want you to recall the details of the study conducted by de Groot (the first paragraph in the passage). Include all the aspects that you can remember about de Groot's study.

Subjects were given an unlimited amount of time to write down what they remembered.

Results

Subjects' written responses were examined to see if the errors observed in the literature were replicated. Other errors were not scored. The results of the experiment are presented in Table 3.

Data from the historically inaccurate passage

The majority of subjects who received the passage from Chase and Simon (1973b) reported that de Groot conducted the random control condition and that he had

Table 3. *Experiment 2: Recall of Chase and Simon (1973b) and Simon and Chase (1973) by undergraduate subjects*

Historically incorrect passage (Chase & Simon, 1973b)

	Percentages[a]		
Type of response	Correct	Error	Omit and other
Random control[b]	6	76	18
5 s exposure[b]	0	71	29
Novices	35	65	0
Masters perfect	41	59	0

[a]$n = 17$.
[b]These responses were not errors for these subjects because they were already present or implied in the text that subjects were asked to read.

Historically correct passage (Simon & Chase, 1973)

	Percentages[a]		
Type of response	Correct	Error	Omit and other
Random control	41	47	12
5 s exposure	41	0	59
Novices	12	82	6
Masters perfect	100	0	0

[a]$n = 17$.

adopted exposure times of 5 s. While these facts are historically incorrect, they were already present in the passage that subjects read. Thus, subjects were merely repeating the errors that appeared in the text. The more interesting data in this condition are those concerning the control group's level of expertise and the masters' performance.

The text stated that the masters' recall was "almost perfect", yet over half of the subjects incorrectly indicated that the masters exhibited perfect recall. In some of the recalls the distortion was very explicit (e.g., "a master chess player could remember exactly where all of the pieces were") and in others it was more implicit (e.g., "the master players were able to reconstruct the position of the chess pieces after viewing them for 5 s"). This seems like a good example of a schema-based change which distorts the original text in the direction of an idealized scientific experiment.

Over half of the subjects incorrectly reported that de Groot's experiment included a novice control group. Chase and Simon (1973b) referred to the less skilled group as "below the master level" and as "[the] weaker players". They intended these phrases to indicate chess players of "expert" and "class" skill. However, we now think that our undergraduate subjects interpreted these phrases to mean "novices" when they read the original text (see Experiment 3).

Data from the historically correct passage

Almost half of the subjects who received the passage from Simon and Chase (1973) reported that de Groot had conducted the random control experiment, despite the fact that the passage they had read clearly indicated that this control condition was conducted by Chase and Simon, and not de Groot. Furthermore, over 80% of the subjects incorrectly reported that de Groot's experiment included a novice control group. While the text explicitly states that the novice data came from Chase and Simon's own work, the information is given in a parenthetical comment and we think many of the subjects may have made this error at the time of comprehension.

No subjects in this condition stated that the masters' recall was perfect. This finding shows how information in the text can help prevent schema distortions. The Simon and Chase (1973) text stated that masters were able to reproduce the positions with "almost perfect accuracy", which is essentially identical to the text used in Chase and Simon (1973b); however, the Simon and Chase (1973) text *also* includes the fact that the masters were "about 93% correct" and this additional information apparently eliminated the distortion that the masters' performance was perfect. The data on this issue are clear for the half-hour recall interval used in the present experiment; however, we suspect that over longer time intervals the Simon and Chase subjects would forget the detailed information about the masters' 93% accuracy and many would make the error of recalling the masters' performance as perfect. The mistake of attributing a 5 s exposure time to de Groot's study, which was regularly observed in the literature, was not replicated in the recall data.

EXPERIMENT 3

Our analysis of the results of Experiment 2 suggested that some of the apparently reconstructive errors might be occurring at the time of comprehension. Therefore we carried out one additional experiment to explore this issue. We have claimed that the Simon and Chase (1973) text is essentially accurate on all of the issues we have been examining in these studies (the reader might be advised to examine the text given in Table 1 to see if they agree). However, given the size of some of our distortions in recall, we thought it was important to see if our undergraduate subjects were interpreting the passage the same way we were.

Method

Seventeen undergraduate subjects from the University of Illinois were presented with the same historically accurate passage (Simon & Chase, 1973) that was given

to half of the subjects in Experiment 2. The instructions were the same as before, except that in place of the section that indicated they were not going to be allowed to look back at the text they were told:

> We would like you to read this text very carefully. Previous research has shown that undergraduates have trouble in understanding this text. After you have read the text you will be asked to answer several questions about the research described in the passage. Please refer back to the text while answering the questions to ensure the accuracy of your response.

Subjects were then presented with a single page of text. The passage to be read (see Table 1) was printed on the top of the page, and the following questions were presented at the bottom of the page:

(1) Did de Groot conduct an experiment with random chess positions?
(2) What were the levels of expertise of the subjects in the recall study conducted by de Groot?
(3) What were the exposure times adopted by de Groot?
(4) What were the recall scores of the grand masters and masters in de Groot's study?

Results

The basic results are given in Table 4. In this comprehension test with the text in front of them none of the subjects made errors on the exposure time or stated that the masters' performance was perfect. Few subjects made the error of attributing the random control to de Groot. (The 47% error rate for the memory subjects in Experiment 2 is reliably higher than the 18% error rate found in the comprehension subjects, $\chi^2 = 4.50$, $p < .05$.) However, most of the comprehension subjects did make the error of stating that de Groot used "novice" subjects. Apparently most undergraduates do not understand the parenthetical phrase "novice players (from our own experiments)" in the way that Simon and Chase intended it. We leave open the issue of whether the authors who have made this

Table 4. *Experiment 3: Comprehension of Simon and Chase (1973), by undergraduate subjects*

	Percentages[a]		
Type of response	Correct	Error	Omit and other
Random control	82	18	0
5 s exposure	100	0	0
Novices	12	88	0
Masters perfect	100	0	0

[a] $n = 17$.

same error in the published literature made it at the time of comprehension as did our undergraduate subjects or if they comprehended it correctly, but recalled it incorrectly.

GENERAL DISCUSSION OF EXPERIMENTS 2 AND 3

The results of these experiments are consistent with the reconstructive remembering hypothesis. Many of the subjects who were presented with a correct description of de Groot's experiment and Chase and Simon's follow-up work incorrectly attributed the random control condition to de Groot, thereby replicating the misrepresentation found repeatedly in the literature and in our experiment on research scientists. Most of these subjects also reported that de Groot's study included a novice control group, thereby replicating another of the citation errors that was found repeatedly in the literature. However, this error appears to be predominantly made at comprehension.

The historically inaccurate passage also resulted in some additional errors that replicate those observed in the literature. Subjects tended to convert the "almost perfect" performance of the masters into "perfect" performance (the strongest finding that could be obtained to support the theory of expert chess memory).

The fourth frequently observed citation error, that the exposure times adopted by de Groot were 5 s, was observed only for recall of the incorrect passage. We think that the published citations of the incorrect exposure time is an interference effect that results from a confusion between the exposure times used and reported in the different studies (e.g., Chase & Simon, 1973a, 1973b; de Groot, 1965, 1966; Jongman, 1968).

While the findings from these experiments are already fairly impressive, there are several reasons to think that they may be underestimates of the strength of the schema-based memory processes. First, the schema-based processes can work only if the reader has the appropriate schema and understands the text. The de Groot study seems fairly easy for undergraduates to understand. Examination of the recalls shows that the great majority of our subjects understood the point of this experiment; however, the recalls also suggest that there was a small group who apparently did not understand the point and therefore could not have given strong schema-based reconstructions. Thus, the strongest schema effects with this type of material would probably be obtained with a subject population of research psychologists who had never heard of the de Groot experiment.

The second factor that may have reduced some of the schema effects was the time interval used in this experiment. In the case of citation errors the interval between reading a text and writing the description of it may be over periods of years or even decades. We think that the effect of time interval may be responsible in the present study for the difference in number of subjects recalling

that the masters showed perfect performance by the group that read the Chase and Simon (1973b) text (59%) and by the group that read the Simon and Chase (1973) text (0%). The only major difference between the texts on this topic was that the Simon and Chase text added an explicit number (93%) for the level of performance of the masters. Most of the subjects who read the Simon and Chase text (76%) included this number in their recalls and thus did not make the schema-based error of recalling that the performance of the masters was perfect. It seems to us that with a longer retention interval most of these subjects would not retain the detailed information about the performance of the masters and so some portion of these subjects would also come to recall the performance as perfect.

It is important to reiterate that the reconstructive remembering hypothesis is not intended to be the only explanation for all of the errors observed in the literature. Clearly, many authors obtained their knowledge of the de Groot experiment from the erroneous Chase and Simon (1973a, 1973b) descriptions. However, as mentioned earlier, this factor alone cannot explain all of the findings since several of the errors found in subsequent citations go beyond the original misrepresentation found in Chase and Simon, and in at least one case it is clear that the author did read at least part of de Groot, but nevertheless committed two citation errors. The reconstructive remembering hypothesis is intended as a complementary explanation of the citation errors that can plausibly account for these additional pieces of evidence.

OTHER CASES OF DISTORTION

Up until this point, we have examined one particular case in great detail. The evidence presented supports the contention that the misrepresentations arose from a process of reconstructive remembering. An obvious question is: how prevalent are such cases of distortion in the scientific literature? The distortions of the de Groot experiments could be just a unique instance. If, on the other hand, this case is part of a more general pattern, then the phenomenon of reconstructive remembering in science may be an important one, with significant implications for the practice of science. In this section, a number of cases similar to the de Groot case are reviewed.

Watson and Rayner (1920)

Watson and Rayner's (1920) study of Little Albert is perhaps the best-known example of misrepresentation in the psychological literature (cf. Cornwell & Hobbs, 1976; Harris, 1979; Prytula, Oster, & Davis, 1976). Like the de Groot

study and the other experiments to be described below, the case of Little Albert is well known to psychologists. In fact, Harris (1979) reports that the study is one of the most widely cited experiments in psychology textbooks. The question addressed by Watson and Rayner's (1920) study was whether it was possible to condition various types of emotional response. They carried out their experiment with an infant, Albert B.

The numerous and varied misrepresentations of this experiment have been well documented, most notably by Harris (1979). The errors include misrepresentations of both the methods and the results. In trying to come up with an explanation for these mis-citations, Harris (1979) observed that some of the errors seem to arise from the fact that authors have a tendency to make experimental evidence conform with their favorite theories. In support of this claim, Harris points out that two groups of scientists each misrepresented the Little Albert experiment in different ways, such that the distorted accounts were more in line with their respective theoretical suppositions. Thus, in the early 1960s, behavior therapists incorrectly claimed that Little Albert developed a rat phobia, and then ten years later, proponents of preparedness theory incorrectly stated, among other things, that Watson himself did not become an aversive stimulus to Albert. This finding can be interpreted in terms of schema-based reconstructive remembering, since the information that was recalled from the experiment seems to have been distorted by the theoretical views of the respective scientists.

In addition, some authors have stated that Little Albert was desensitized after the experiment, a standard procedure in experiments of this sort, when in fact he was not. This error provides an example of authors adopting an idealized experiment schema to reconstruct the details of an original experiment which did not conform to the idealized schema. Thus, they describe the experiment the way in which it should have been done, not the way in which it was done. Note that this distortion cannot be accounted for by biases resulting from an individual's specific theory. Rather, it is related to the general way in which experiments of this type tend to be conducted. This observation provides further support for the schema-based reconstruction hypothesis.

Allport and Postman (1945)

Allport and Postman (1945) conducted a classic study investigating the psychology of rumor that was very similar to Bartlett's (1932) experiments with the method of serial reproduction. In the Allport and Postman experiments, a slide was projected onto a screen while six or seven subjects, who had not seen the picture, waited in a room nearby. The first subject entered and took a position where he or she could not see the screen. Someone in the audience then described

the picture to the subject. A second subject entered the room and stood beside the first subject, who proceeded to tell the second subject everything that the original subject could remember about what he or she had been told was in the picture. A third subject entered to hear the story from the second subject, and the procedure was repeated several times. The experimenters found that the subjects' reports often became increasingly distorted as the message was passed along to each successive subject.

It is indeed ironic that this study of the psychology of rumor has also been subject to misrepresentation. Treadway and McCloskey (1987) have pointed out that both the methods and the results of the Allport and Postman study have been repeatedly distorted in the psychological literature. Furthermore, several types of citation errors can also be found in the legal literature on eyewitness testimony, apparently because legal scholars relied on the erroneous secondary citations in the psychological literature. The most blatant error, committed in the psychological and legal literature and in the courtroom as well, was the claim that the subjects who were responsible for the distortions had viewed the picture. In fact, the subjects' reports were based on what they had heard from others, not on what they had seen.

These distortions of the Allport and Postman study seem to be another example of schema-based reconstructive remembering. The experiment, as described in both legal and psychological secondary sources, seems to have been distorted by an idealized experiment schema based on the theoretical belief that expectations can affect direct visual perception. The actual Allport and Postman study cannot support such a claim since the subjects in that study were not eyewitnesses. Once again it appears that researchers have inadvertently distorted the details of an original study to be more consistent with their schema.

Moray (1959)

Moray (1959) conducted an early study of selective attention which, as Loftus (1974) has pointed out, has also been subsequently misrepresented in the literature. Moray's experiment consisted of a dichotic listening task where subjects were required to repeat out loud a message presented to one ear while other material was being presented in the other ear. Moray discovered that the material presented in the ear that is to be rejected can sometimes break through the attentional barrier. In particular, when the subject's name is presented in the unattended channel and when no prewarnings are given, 33% of the subjects reported hearing their name. This experimental finding has become widely known as "the cocktail party phenomenon".

Loftus (1974) cites several authors who incorrectly claim that Moray found that the subject's name is detected *every* time it is presented in the unattended ear. In

terms of the taxonomy presented earlier, this is a results error since an experimental finding has been cited incorrectly. The mistake seems to be a typical example of schema-based reconstructive remembering, where the strongest possible result to support the theory (100%) is given in place of the actual result (33%).

Walster (1965)

Walster (1965) conducted an important study of the effects of self-esteem on romantic liking that, according to Rubin (1974), is frequently cited in social psychology textbooks. All of the subjects in this experiment were women who were 18 or 19 years old. A short time after the subject arrived for the experiment, a male confederate entered the same room. The confederate struck up a conversation with the subject, and during that conversation he conveyed personal interest in the subject, and eventually asked her out on a date. Soon after the date was made, the experimenter entered the room, and informed the subject that she would be given a psychological test. The experimenter asked the confederate to take the place of a co-interviewer who supposedly had not yet appeared, and give the subject the test. Once this test was completed, the experimenter handed the subject either an extremely flattering or an extremely disparaging analysis of a personality test she had taken a few weeks earlier. The experimenter then had the subject fill out a questionnaire asking about her feelings concerning four people, one of whom was the male confederate. The results of the experiment revealed that women whose self-esteem had been temporarily lowered liked the confederate significantly more than did women whose self-esteem had been temporarily raised.

Rubin (1974) has pointed out that several authors citing Walster's (1965) experiment have incorrectly reported the procedure adopted. In the published experiment receipt of approval from the male confederate preceded the self-esteem manipulation but several authors have reversed the order of these two events. Rubin (1974) proposes that one of the primary explanations of this error is that the re-ordered sequence seems to provide a better fit to social psychologists' theoretical frameworks than the actual sequence does. So again we find that the details of a well-known experiment are reconstructed based on a schema of a canonical experiment supporting a specific theoretical point, thereby leading to predictable distortions. Rubin (1974) made a similar observation, comparing the distortions of the Walster experiment with the findings of Allport and Postman (1945).

Schachter (1951)

Schachter (1951) conducted an important investigation of "deviation, rejection, and communication" that, according to Berkowitz (1971), is one of the most

widely cited experiments in social psychology. In this study, the subjects' degree of attraction to their experimental groups was manipulated by assigning them to clubs whose activities were either more or less preferred. Each club then proceeded to discuss the case history of a juvenile delinquent. The relevance of the discussion topic to each group was also manipulated. In each club there were three paid participants, one of whom persisted with a deviant opinion throughout the discussion, another who maintained a modal opinion, and a third who began the discussion with a deviate opinion but gradually moved toward the modal view. The primary question of interest was the amount of communication directed towards each of the three paid participants during the discussion and the subjects' attitudes towards these confederates at the end of the session. The results of the experiment are complex and difficult to interpret since the results obtained from various dependent variables were not consistent.

Berkowitz (1971) has pointed out that the results of Schachter's (1951) well-known experiment have been misrepresented by several authors. In his analysis of the citation errors, Berkowitz observed that some authors have misrepresented some of the experimental findings to assimilate them to their own expectations. Furthermore, he also found that textbook accounts of Schachter's (1951) experiment have omitted several important qualifications, or sometimes have even added details.

The picture that emerges is that authors citing the Schachter (1951) experiment have transformed a complex and somewhat inconsistent set of findings into a clean experimental result, just as one would expect if researchers were reconstructing their recall of the experiment from an idealized experiment schema. Berkowitz himself recognized this connection, citing Bartlett (1932) and Allport and Postman (1945). Once again, the overwhelming tendency to remember a simpler, more consistent depiction of a well-known experiment has led to repeated citation errors.

DISCUSSION

This review of other cases of misrepresentations of experiments in the psychological literature has revealed a common set of characteristics. Like the de Groot study, each of the experiments described above is an important and widely cited work. The distortions seem to occur as the famous findings are disseminated to a larger audience. It is at this point that the process of reconstructive remembering tends to exert its influence. While we have no laboratory evidence to support our position that these cases were caused by the properties of human memory, the pattern of errors that is exhibited across the various cases is uniform and consistent with the "signature" of reconstructive remembering. In all of the cases reviewed, the recall of scientists seems to be driven by an idealized experiment schema that distorts certain details of the actual experiment being reported in

predictable ways. Several of the authors who wrote papers to set the historical record straight on these cases of misrepresentation have commented on this fact, pointing out the similarities with the findings of Bartlett (1932) and Allport and Postman (1945).

Why is this pattern of behavior exhibited by scientists? First, scientists frequently prefer to cite the original study in an area but, because it is the first of its kind, such initial experiments often have flaws which do not fit within the canonical experiment schema. With hindsight, it is easy to describe the experiment as it should have been conducted, rather than as it actually was conducted. The purported desensitization of Little Albert is a good example. Second, in thinking about the results of a particular experiment, it is easier to keep in mind a simple and coherent generalization than it is to try and remember the numerous, and sometimes incoherent, details of the experiment. As Berkowitz (1971) noted, having a consistent and economic summary of an experiment allows us to convince ourselves that we have indeed apprehended the major findings and implications of the study. In addition, when describing the experiment to others, whether it is in an article or in a classroom, a simple generalization allows one to be clearer and more succinct. While it is usually possible to be both concise and accurate, simplification sometimes has its perils. As the various cases reviewed above indicate, describing a study from memory can lead to significant distortions.

It is difficult to establish the frequency with which such cases of misrepresentation occur. The cases described above were identified primarily through our examination of the literature and from colleagues who had heard of instances of mis-citation similar to that of de Groot's. There does not seem to be any other systematic way of identifying such cases because the articles that appear documenting the misrepresentations tend to be buried in specialized journals. That all of the cases identified above were from the psychological literature is probably due to the fact that this is the discipline with which we and our colleagues are most familiar. It seems highly unlikely that such misrepresentations are committed only by psychologists, as evidenced by the examples from other disciplines cited in the introduction (cf. also Brush, 1989). Consequently, there is a reason to believe that other similar cases of misrepresentation of scientific fact exist, some already documented and others waiting to be discovered. In fact, we would appreciate it if researchers would send us other examples of distortions such as these.

CONCLUSIONS

The investigation described in this paper represents a new contribution to the understanding of human memory and philosophy of science. As far as we know, it is the only theoretical and experimental study of the impact of memory on the

process of doing science. The subject of the investigation was Adriaan de Groot's (1946, 1965) widely cited experiment on the relationship between chess expertise and memory recall. This important study of expertise has had a prominent impact on the literature, as evidenced by the exceptionally large number of citations collected. However, an examination of citations of de Groot's study reveals that many authors have misrepresented de Groot's work. While the gist of the findings has been consistently conserved, a regular pattern of citation errors is observed. Furthermore, the severity of the errors has increased with time, primarily in review papers, textbooks, and popular works. Our experiment on research psychologists' recall of results of de Groot's experiment reveals that the misconceptions regarding de Groot's memory experiment are indeed widespread. While some of these distortions undoubtedly resulted from authors who relied only on Chase and Simon's account for describing de Groot's work, this explanation cannot account for all of the evidence reviewed above.

We hypothesized that some of these distortions can be attributed to reconstructive memory processes of various types, including source confusions, interference effects, and schema-based reconstructions. With regard to the latter, scientists seem to have a schema for the canonical scientific experiment which consists of a clean test of a particular theory with appropriate controls and clear, unambiguous results. At the time of recall, there will be a strong tendency for the idealized experiment schema to override the actual details of a given experiment. Two memory experiments were conducted to explore this hypothesis. Subjects replicated the citation errors that were consistently observed in the literature, thereby providing support for the reconstructive remembering hypothesis. A review of other cases of distortion of well-known psychological experiments shows that the case study analyzed in detail in this paper is by no means an isolated one. Furthermore, the errors made in these other cases are similar to those observed in the de Groot case in that they too exhibit the signature of schema-based reconstructive remembering; this suggests that reconstructive memory effects in reporting of scientific experiments may be more prominent than one might have expected. These findings suggest, perhaps surprisingly, that remembering in science may not be all that different from remembering in daily life.

However, the implications of the fallibility of remembering within the context of scientific endeavor are much more severe than in everyday life. First, problems can occur when one's theoretical bias leads to misrepresentations when citing the details of the original study. For instance, in the case of Little Albert, two groups of scientists both misrepresented the same experiment in slightly different ways, such that the distorted accounts were more in line with their respective theoretical suppositions. The same tendency was observed in several other cases reviewed in the previous section. This example clearly shows that scientists may sometimes be incorrect, even at the level of gist. Stated more generally, memory for gist is a relative notion that results in *coherence* between a set of observations and a

particular point of view (e.g., a theoretical framework), but does not necessarily ensure a strong *correspondence* between those observations and the events that actually transpired. Second, it is not at all clear which facts should only be considered minor details and which are actually relevant at the level of gist. A good example is the set of distortions of Moray's dichotic listening experiment. Whether subjects hear their name on 33% of the trials or on every trial may seem to be a trivial detail. In fact, this distinction is very important since the two results put very different constraints on a viable theory of attention. Thus, reconstructive remembering in science can have significant practical implications.

Our study shows that the normal operation of human memory leads to errors in the scientific enterprise that go against the scientific norm requiring accuracy in reporting the work of other scientists. In fact, certain characteristics of the institution of science (e.g., lab notebooks, published journals, replication of experiments) can be viewed as tools designed to reduce the memory errors of individual scientists. Our study suggests that scientists also need to be aware of the power of schema-based memory processes so that they will not rely on memory as a replacement for careful re-reading of original sources. In fact, we have recently found an example of a writer attempting to make his readers aware of this problem in an introductory psychology text written in 1939! In that text, Lawrence Cole reviewed Bartlett's (1932) work on reconstructive remembering and then, in a prophetic aside to his readers, states that Bartlett's research "might well cause the student of psychology to do some reflecting, since the very textbook which he reads is a reproduction often many times removed from that source (human behavior) to which it professedly refers" (Cole, 1939, p. 516).

To conclude, the case study presented here serves as a humbling reminder that scientists, like all other human beings, are fallible. With few exceptions, people do not have the capability to flawlessly "store" the past and "retrieve" it at will. As Henri Bergson pointed out many years ago, "The brain's function is to choose from the past, to diminish it, to simplify it, to utilize it, but not to preserve it" (cited in Rosenfield, 1988, p. 12).

References

Allport, G.W., & Postman, L.J. (1945). The basic psychology of rumor. *Transactions of the New York Academy of Sciences*, *2*(8), 61–81.

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

Berkowitz, L. (1971). Reporting an experiment: A case study in leveling, sharpening, and assimilation. *Journal of Experimental Social Psychology*, *7*, 237–243.

Brewer, W.F., & Nakamura, G.V. (1984). The nature and functions of schemas. In R.S. Wyer Jr. and T.K. Srull (Eds.), *Handbook of social cognition* (Vol. 1, pp. 119–160). Hillsdale, NJ: Erlbaum.

Brush, S.G. (1989). Prediction and theory evaluation: The case of light bending. *Science*, *246*, 1124–1129.

Chase, W.G., & Simon, H.A. (1973a). The mind's eye in chess. In W.G. Chase (Ed.), *Visual information processing* (pp. 215–281). New York: Academic Press.

Chase, W.G., & Simon, H.A. (1973b). Perception in chess. *Cognitive Psychology*, *4*, 55–81.

Cole, L.E. (1939). *General psychology*. New York: McGraw-Hill.

Cooper, L. (1935). *Aristotle, Galileo and the Tower of Pisa*. Ithaca, NY: Cornell University Press.

Cornwell, D., & Hobbs, S. (1976). The strange saga of little Albert. *New Society*, *35*, 602–604.

de Groot, A.D. (1946). *Het denken van den schaker* [Thought and choice in chess]. Amsterdam: North-Holland.

de Groot, A.D. (1965). *Thought and choice in chess*. The Hague: Mouton.

de Groot, A.D. (1966). Perception and memory versus thought: Some old ideas and recent findings. In B. Kleinmuntz (Ed.), *Problem solving: Research, method, and theory* (pp. 19–50). New York: Wiley.

Djakow, I.N., Petrowski, N.W., & Rudik, P.A. (1927). *Psychologie des Schachspiels* [The psychology of chess]. Berlin: de Gruyter.

Ellenberger, H.F. (1972). The story of "Anna O": A critical review with new data. *Journal of the History of the Behavioral Sciences*, *8*, 267–279.

Engle, R.W., & Bukstel, L. (1978). Memory processes among bridge players of differing expertise. *American Journal of Psychology*, *91*, 673–689.

Gilhooly, K.J. (1988). *Thinking: Directed, undirected and creative* (*2nd edn.*). London: Academic Press.

Gillespie, R. (1988). The Hawthorne experiments and the politics of experimentation. In J.G. Morawski (Ed.), *The rise of experimentation in American psychology* (pp. 114–137). New Haven, CT: Yale.

Haines, H., & Vaughan, G.M. (1979). Was 1898 a "great date" in the history of experimental social psychology? *Journal of the History of the Behavioral Sciences*, *15*, 323–332.

Harris, B. (1979). Whatever happened to Little Albert? *American Psychologist*, *34*, 151–160.

Hayes, J.R. (1978). *Cognitive psychology: Thinking and creating*. Homewood, IL: Dorsey Press.

Hirsch, E.D., Jr. (1988). *Cultural literacy: What every American needs to know*. New York: Vintage Books.

Hofstadter, D.R. (1979). *Gödel, Escher, Bach: An eternal golden braid*. New York: Basic Books.

Holton, G. (1988). Einstein, Michelson, and the "crucial" experiment. In G. Holton (Ed.), *Thematic origins of scientific thought: From Kepler to Einstein* (rev. edn. pp. 279–370). Cambridge, MA: Harvard University Press.

Johnson-Laird, P.N., & Wason, P.C. (1977). *Thinking: Readings in cognitive science*. Cambridge: Cambridge University Press.

Jongman, R.W. (1968). *Het oog van de meester* [The eye of the master]. Amsterdam: van Gorcum.

Loftus, E.F. (1974). On reading the fine print. *Quarterly Journal of Experimental Psychology*, *27*, 324.

Loftus, E.F., Miller, D.G., & Burns, H.J. (1978). Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology: Human Learning and Memory*, *4*, 19–31.

Martin, L. (1986). "Eskimo words for snow": A case study in the genesis and decay of an anthropological example. *American Anthropologist*, *88*, 418–423.

McCloskey, M., & Zaragoza, M. (1985). Misleading postevent information and memory for events: Arguments and evidence against memory impairment hypotheses. *Journal of Experimental Psychology: General*, *114*, 1–16.

Merton, R.K. (1968). The Matthew effect in science. *Science*, *159*, 56–63.

Merton, R.K. (1988). The Matthew effect in science, II: Cumulative advantage and the symbolism of intellectual property. *Isis*, *79*, 606–623.

Moates, D.R., & Schumacher, G.M. (1980). *An introduction to cognitive psychology*. Belmont, CA: Wadsworth.

Moray, N. (1959). Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology*, *11*, 56–60.

Neisser, U. (1981). John Dean's memory: A case study, *Cognition*, *9*, 1–22.

Neisser, U. (1982). *Memory observed: Remembering in natural contexts*. San Francisco: Freeman.

Neisser, U., & Winograd, E. (1988). *Remembering reconsidered: Ecological and traditional approaches to the study of memory*. Cambridge: Cambridge University Press.

Posner, M.I. (1988). Introduction: What is it to be an expert? In M.T.H. Chi, R. Glaser, & M.J. Farr (Eds.), *The nature of expertise* (pp. xxix–xxxvi). Hillsdale, NJ: Lawrence Erlbaum.

Price, D.J. (1986). *Little science, big science . . . and beyond*. New York: Columbia University Press.

Prytula, R.E., Oster, G.D., & Davis, S.F. (1976). The "rat rabbit" problem: What did John B. Watson really do? *Teaching of Psychology*, *4*, 44–46.

Pullum, G.K. (1989). The great Eskimo vocabulary hoax. *Natural Language and Linguistic Theory*, *7*, 275–281.

Rosenfield, I. (1988). *The invention of memory: A new view of the brain*. New York: Basic Books.

Rothman, T. (1989). *Science à la mode: Physical fashions and fictions*. Princeton, NJ: Princeton University Press.

Rubin, Z. (1974). A new order of things: A case study in the reconstruction of an experimental procedure. *Representative Research in Social Psychology*, *5*, 81–87.

Saariluoma, P. (1985). Chess players' intake of task-relevant cues. *Memory and Cognition*, *13*, 385–391.

Sanford, A.J. (1985). *Cognition and cognitive psychology*. New York: Basic Books.

Schachter, S. (1951). Deviation, rejection, and communication. *Journal of Abnormal and Social Psychology*, *46*, 190–207.

Schacter, D.L. (1986). The psychology of memory. In J.E. LeDoux & W. Hirst (Eds.), *Mind and brain: Dialogues in cognitive neuroscience* (pp. 189–214). Cambridge: Cambridge University Press.

Schacter, D.L., Harbluk, J.L., and McLachlan, D.R. (1984). Retrieval without recollection: An experimental analysis of source amnesia. *Journal of Verbal Learning and Verbal Behavior*, *23*, 593–611.

Seamon, J.G. (1980). *Memory and cognition: An introduction*. New York: Oxford University Press.

Simon, H.A., & Chase, W.G. (1973). Skill in chess. *American Scientist*, *61*, 394–403.

Treadway, M., & McCloskey, M. (1987). Cite unseen: Distortions of the Allport and Postman rumor study in the eyewitness testimony literature. *Law and Human Behavior*, *11*, 19–25.

VanLehn, K. (1989). Problem solving and cognitive skill acquisition. In M.I. Posner (Ed.), *Foundations of cognitive science* (pp. 527–579). Cambridge, MA: MIT Press.

Vicente, K.J. (1988). On adapting the memory recall paradigm to evaluate interfaces. *Acta Psychologica*, *69*, 249–278.

Vicente, K.J., & de Groot, A.D. (1990). The memory recall paradigm: Setting the historical record straight. *American Psychologist*, *45*, 285–287.

Walster, E. (1965). The effect of self-esteem on romantic liking. *Journal of Experimental Social Psychology*, *1*, 184–197.

Watson, J.B., & Rayner, R. (1920). Conditioned emotional reactions. *Journal of Experimental Psychology*, *3*, 1–15.

Whorf, B.L. (1940). Science and linguistics. *Technology Review*, *42*, 229–231, 247–248. Reprinted in J.B. Carroll (Ed.) (1956). *Language, thought, and reality: Selected writings of Benjamin Lee Whorf*. Cambridge: MIT Press.