H.M.'s personal crossword puzzles: Understanding memory and language

Brian G. Skotko, David C. Rubin & Larry A. Tupler

Duke University, Durham, NC, USA
Duke University Medical Center, Durham, NC, USA

Published online: 19 Feb 2008.

To cite this article: Brian G. Skotko , David C. Rubin & Larry A. Tupler (2008) H.M.'s personal crossword puzzles: Understanding memory and language, Memory, 16:2, 89-96, DOI: 10.1080/09658210701864580

To link to this article: http://dx.doi.org/10.1080/09658210701864580

PLEASE SCROLL DOWN FOR ARTICLE
H.M.’s personal crossword puzzles: Understanding memory and language

Brian G. Skotko and David C. Rubin
Duke University, Durham, NC, USA

Larry A. Tupler
Duke University Medical Center, Durham, NC, USA

The amnesic patient H.M. has been solving crossword puzzles nearly all his life. Here, we analysed the linguistic content of 277 of H.M.’s crossword-puzzle solutions. H.M. did not have any unusual difficulties with the orthographic and grammatical components inherent to the puzzles. He exhibited few spelling errors, responded with appropriate parts of speech, and provided answers that were, at times, more convincing to observers than those supplied by the answer keys. These results suggest that H.M.’s lexical word-retrieval skills remain fluid despite his profound anterograde amnesia. Once acquired, the maintenance of written language comprehension and production does not seem to require intact medial temporal lobe structures.

Since 1953, patient H.M. has been shaping our understanding of human memory. At the age of 27 he underwent a medial temporal lobe (MTL) resection for relief of intractable epilepsy, resulting in a massive anterograde amnesia (Scoville, 1954, 1968; Scoville, Dunsmore, Liberson, Henry, & Pepe, 1953; Scoville & Milner, 1957). He has been the subject of countless articles, books, and lectures, but one topic has scarcely been mentioned: H.M. loves to work on crossword puzzles (Markowitsch, & Pritzel, 1985, p. 205). Despite the fact that he has acquired only limited semantic knowledge since his operation (Corkin, 1984, 2002; Gabrieli, Cohen, & Corkin, 1988; O’Kane, Kensinger, & Corkin, 2004; Skotko et al., 2004), he continues to work on two or more puzzles each day. Those close to him substantiate that he frequently engages in challenging puzzles featured in books published by the New York Times, The Hartford Times, and Merriam Webster. According to his own reports, he began solving crossword puzzles that were printed in his local newspaper around the age of 15, and he has continued with the pastime even today.

Crossword puzzles require a wide range of cognitive faculties. As such, they have provided us with new information about H.M.’s ability to acquire new vocabulary even after the removal of his medial temporal lobes. Skotko et al. (2004) examined H.M.’s performance on a series of specially designed crosswords, which tested his...
preoperative and postoperative semantic knowledge. On these tests, H.M. demonstrated that he was able to acquire new postoperative semantic information when he could anchor it to mental representations established preoperatively. For example, when asked which childhood disease was successfully treated by the Salk vaccine, H.M. eventually responded with “polio”, despite the vaccine being invented 2 years after his operation.

Because crosswords draw on written language production and reading comprehension, including lexical, grammatical, and spelling skills, they can also offer a rich insight into a person’s linguistic skills. Hambrick, Salthouse, and Meinz (1999) had adults, ranging in age from 18 to 80, complete a series of skills tests that they believed would be associated with crossword-puzzle proficiency. Correlation analysis revealed that the three major predictors of crossword success were general knowledge, previous crossword puzzle experience, and word-retrieval skills. Underwood, Diehim, and Batt (1994) found that expert crossword solvers’ puzzle proficiency was predicted by word generation from a string of letters, performance on anagrams, sensitivity to the suffix of a word, sensitivity to the pseudosuffix of a word, and vocabulary scores.

Many studies of H.M.’s language ability have been conducted based on analyses of his speech, transcripts of his speech (which lose much of the intonation and context), standardised tests, and tests developed specifically to probe H.M.’s abilities (Kensinger, Ullman, & Corkin, 2001; MacKay, Burke, & Stewart, 1998a; MacKay & James, 2001; MacKay, Stewart, & Burke, 1998b; Skotko, Andrews, & Einstein, 2005). Together, these studies provide a detailed and sometimes contentious view (MacKay, 2006) of how bilateral MTL damage affects language. This study does not intend to settle such issues, but it does contribute a unique form of evidence. By studying H.M.’s leisure-time non-cryptic puzzle books, we deliberatively bypassed laboratory tasks and investigated his language skills using a more ecologically valid approach—one in which time constraints and any hint of a test were removed. Here we ask: What can we learn about H.M.’s lexical access from his personal crosswords? How robust are his language skills in a self-chosen task carried out without pressure? From performance under these circumstances, what can we conclude about the role of MTL structures in the maintenance of language comprehension and production?

**METHOD**

**Participant**

During this 3-year study, H.M. was 72–74 years old. His experimental resection of MTL structures at age 27 included all MTL structures, except approximately the caudal 2 cm of the hippocampus and parahippocampal gyrus (Corkin, Amaral, González, Johnson, & Hyman, 1997). H.M. completed 12 years of education at a technical high school.

**Materials**

We collected copies of 277 puzzles from six books that H.M. had worked on during his leisure time between 1997 and 1999. These books included the *Pocket Patter* of Dell Puzzle Magazines (January, 1998), the *Penny Press Means Puzzle Pleasure* (January, 1998), and *Large Print Crosswords*. The identification of the remaining three books could not be ascertained from the photocopies, but their difficulty level was comparable to that of the other puzzle books. The 277 puzzles contained 2834 clues in which H.M. produced an error, as coded below.

**Coding of responses**

Based on observations at the MIT Clinical Research Center and at his residence, H.M. has never been known to seek assistance in completing his puzzles, either from others or from the answer key. Despite this, however, we adopted a conservative approach in defining which answers were truly his. We purposely ignored all of his correct responses and instead coded three types of error: (1) misspellings, (2) alternative responses (instances in which his response differed from that of the answer key), and (3) omissions (instances in which he provided no response). By coding his errors only, we have excluded the possibility that he might have used the answer key for these particular clues.

We defined “misspellings” as cases in which H.M.’s answer was the same as that of the answer key, except for one or two letters (e.g., “penda-lum” for “pendulum”). In cases where he misspelled a word that differed from the answer key, we did not code the response as a misspelling but
instead coded it as an “alternative response” (e.g., for the clue “____ Rica”, the correct answer was “Costa”. H.M. wrote “porta”, which is an apparent misspelling of “Puerto”). In the case of “omissions”, H.M. may not have responded to a clue because (1) he did not have sufficient time to reach the clue while working on a puzzle, which is not of theoretical interest, or (2) he did have ample time to complete the clue but could not, which is of interest. As noted in more detail at the beginning of the Results, H.M. tends to do all the Across cues in order, followed by all the Down cues in order. We therefore inferred that he had ample opportunity to complete a given clue if (1) all of the spaces surrounding the clue on the crossword grid were either black boxes or boxes filled in with a letter or (2) the clues immediately before and after the unanswered clue had been answered. Thus, for the first clue to be coded as an omission, only the second clue had to be answered; likewise, for the last clue to be coded as an omission, only the penultimate clue had to be answered. These strict criteria for defining an omission reasonably suggested that H.M. had read the clue but could not think of an answer.

We also considered the possibility that when H.M. encountered a clue, a portion of the answer was already filled in due to the crisscrossing nature of crossword puzzles. All shared letters were recorded as orthographic cues. (For example, if the words “black” and “felt” crossed, the letter “l” would be recorded as an orthographic cue for both words, as it was not possible for us to determine which clue he had answered first. He could have answered “black” first and benefited from the “l” in solving “felt”, or he could have answered “felt” first and benefited from the “l” in solving “black”.) However, some orthographic cues were wrong and actually hampered his ability to solve a connecting clue correctly. (Suppose in the previous example that H.M. responded with “brown” instead of “black”. The “r” would not help in solving “felt”. Therefore, the “r” was recorded as a false orthographic cue.) All errors that included a false orthographic cue were excluded from our analyses, because we considered H.M. to be disadvantaged by a previous response. It would be inaccurate to say that he did not know the answer “felt” when given false orthographic cues, even of his own creation. Thus we use only the clearest cases for analysis: errors as opposed to correct responses and only errors that are not influenced by another error. Our measures give a conservative estimate of H.M.’s abilities. The only exception to this is that we might slightly underestimate the number of omissions, but this is not a key measure for our inferences about H.M.’s linguistic abilities.

**Data analysis**

In what order did H.M. solve crossword puzzles? To answer this question, we counted the number of errors that came from Across clues and from Down clues.

What types of errors did H.M. make on his puzzles? Here, we counted the number of misspellings, omissions, and alternative responses.

How robust were H.M.’s lexical access skills? To address this issue, we performed three separate analyses of H.M.’s alternative responses. First, using the Merriam Webster’s online dictionary (www.m-w.com), we noted whether his alternative response was a word, pronounceable nonword, or non-pronounceable nonword. Second, we compared the frequency of his responses with that of the answer key using the tagged version of the Standard Corpus of Present-Day American English, a data set of approximately 1,014,000 words sampled from a wide range of styles and varieties of prose (Kučera & Francis, 1982). If his responses were more common in the English language than words in the answer key, frequency values would be greater than those of the answer key. Conversely, if his responses were less frequent than those of the answer key, his frequency values would be lower. Words that did not appear in the corpus were assigned a frequency of zero. Because zero represents a truncated value, we used Wilcoxon signed-ranks tests to measure the difference between H.M.’s word frequencies and those of the answer key, both with and without the zero entries.

In addition, using a two-column format, we randomly mixed H.M.’s alternative responses with those of the answer key for each clue and asked four healthy, highly educated volunteers to select what they thought was the “best answer” given the clues and orthographic cues. These judges included two women (ages 21 and 23) and two men (ages 20 and 21) who were each paid $20.00. Three were in their fourth year of undergraduate studies at Duke University; the other had a BS degree from Duke. All were native English speakers; one of them solved crossword puzzles on a frequent basis.
The judges were blind to the purpose of the experiment and did not know that one of the two answers was from H.M. The score was the percentage of times in which the judges chose H.M.’s answer over that of the answer key. We reasoned that a high percentage would suggest that H.M. generated appropriate responses to difficult clues. A low percentage would suggest that his language facility was insufficient to respond with convincing answers.

How accurate was H.M.’s comprehension of the crossword clues? Here, we analysed the data in the following three ways. First, we counted the number of times that H.M. generated a word for a clue that was a synonym, definition, category, or fill-in-the-blank. “Synonyms” were defined as a one-word clue or a two-word clue in which one of the two words was a conjunction, article, or preposition (e.g., Clue: recognized, Answer Key: knew). “Definitions” were defined as multiple-word clues in which at least two words were not conjunctions, articles, or prepositions (e.g., Clue: search carefully, Answer Key: comb). “Definitions” were generally more specific than “Synonyms” but clearly shared many of the same properties. “Categories” were clues that asked the solver to generate an example from a list or a list from an example (e.g., Clue: nautical word, Answer Key: bow). Clues were always assigned to “Categories” before “Definitions”. “Fill-in-the-blanks” were clues that required completion of a common phrase or proper name (e.g., Clue: red as a ____, Answer Key: beet).

We also counted the number of times H.M. responded with a proper noun or proper adjective when the clue called for one, and we compared the part of speech of his alternative responses with those of the answer key in the context of the given clues. If he had preserved grammar, we would expect him to generate the appropriate part of speech, even if he answered the clue incorrectly.

RESULTS

In what order did H.M. solve crossword puzzles?

Skotko et al. (2004) observed that in a laboratory setting, H.M. typically solved his crossword puzzles in a linear fashion: He began with 1 Across, completed the Across clues, and then completed the Down clues. He did not benefit from the crisscrossing nature of the puzzles while completing the Across clues. Only when he solved the Down clues did he benefit from letters already in place.

It is likely that he used the same pattern for solving puzzles at his residence because he solved more Across clues than Down clues, possibly due to time constraints. We also observed more errors from Across clues (54%) than from Down clues (45%). In addition, H.M.’s errors, especially between clues numbered 10 and 59, were evenly distributed, with a slightly higher percentage for clues numbered 1 through 9, which could reflect the fact that he attempted more smaller-numbered clues. For example, many of the puzzles had only the first 10 or so clues completed. His attention was likely distracted, and he never returned to these puzzles. In short, the order of the clues within the puzzle did not matter. This finding is consistent with a report by Hambrick et al. (1999), who found that correlations between the ordinal position of the clues and the probability of solving them were typically not significant. On average, H.M. completed 45% of each puzzle (SD = 25%).

What types of errors did H.M. make on his puzzles?

Of the 2834 coded errors, 132 of them asked for a foreign word. We excluded these items from the present analysis because H.M. did not regularly speak a language other than English. Of the remaining 2702 clues, 935 had false orthographic cues (i.e., H.M. had supplied an incorrect response for a crisscrossing clue). We excluded these items from the analysis because H.M. would have been misled when primed with the wrong letters. Of the remaining 1767 clues, 48% were alternative responses, 49% were omissions, and only 3% were misspellings. Thus, roughly half of H.M.’s errors were the wrong response, and the other half represented a failure to respond. The fact that his spelling errors were so infrequent suggests that he did not have any unusual spelling difficulties in the context of these puzzles. In fact, his misspellings were often creative deviations of common words to fit the constraints of the crossword puzzles (e.g., Clue: ___ sauce, Answer Key: tartar, H.M.’s Response: tarter; Clue: artist’s need, Answer Key: easel, H.M.’s Response: easel;
How robust were H.M.’s lexical access skills?

For 851 answers, H.M. provided an alternative response. Of these, 43 clues called for an abbreviation as a response (e.g., *Clue*: agents; *Abbrev.*: dep). These items were excluded from the analyses because of the difficulty in assigning a status of word or nonword to an abbreviation. Of the 808 remaining cases, approximately 91% of H.M.’s alternative responses were valid words, according to Merriam Webster’s online dictionary. Only 7% were pronounceable nonwords (e.g., *Clue*: gloomier, *Answer Key*: darker, H.M.’s *Response*: dirier; *Clue*: slightly open, *Answer Key*: ajar, H.M.’s *Response*: agap), and 2% were unpronounceable nonwords (e.g., *Clue*: sacred image, *Answer Key*: icon, H.M.’s *Response*: iool; *Clue*: divers of taxis, *Answer Key*: hacks, H.M.’s *Response*: cabks). Proper names and clear misspellings of valid words were counted as “valid words”.

We next compared the word frequencies of H.M.’s responses with those of the answer key (Kuc`era & Francis, 1982). Of the 732 cases in which his alternative response was a valid word, he provided a word of a higher frequency 99% of the time, with his mean response being 267.0 frequency units higher than the key (*SD* = 2042.3). He provided a word of a lower frequency 216 times, with his mean response being 112.5 frequency units lower than the key (*SD* = 404.0). In 99 cases, his frequency exactly matched that of the answer key. On average, his response was 118.9 (*SD* = 1552.5) frequency units higher than the answer key, suggesting that H.M.’s responses were more common in the English language than answers from the key (H.M.: *M* = 168.5, *SD* = 1541.9; *Answer Key*: *M* = 49.6, *SD* = 298.1; *p* < 0.0001).

We also analysed the word frequencies without the zero frequency values in H.M.’s response and that of the answer key, because they represented truncations. Of the 343 cases in which H.M.’s alternative response was a valid word, there were 203 cases in which he provided a word with a higher frequency than that of the answer key, 127 cases of a lower frequency, and 13 cases of a tie. On average, his response was higher than the answer key, again suggesting that H.M.’s responses were more common in the English language than the answers of the key (H.M.: *M* = 225.9, *SD* = 1635.8; *Answer Key*: *M* = 95.8, *SD* = 414.0; *p* < .0001).

We next asked: How reasonable were H.M.’s answers? We know that he typically responded with a valid word that had, on average, a higher frequency than the answer key, but these words could have been inappropriate and irrelevant for the given clues. To determine how often H.M.’s responses were appropriate, we asked the four judges to select what they thought was the “best answer”, given the clue and orthographic cues, from a two-column list of responses produced by H.M. and the answer key, randomly interchanged. By counting the percentage of times in which these individuals chose H.M.’s response, we identified the lower limit at which his answers were as good as or better than those of the answer key. Of the 851 cases where H.M. provided an alternative response, on average, H.M.’s answer was chosen about one third of the time (*M* = 33%, *SD* = 3%). The one judge who was a frequent puzzle solver (approximately five times each week) chose H.M.’s answers 29% of the time. At least one of the four judges chose H.M.’s response in approximately 56% of the cases; at least two judges chose H.M.’s responses in 39% of the cases; at least three judges in 25% of the cases; and all four judges in 12% of the cases (e.g., *Clue*: ___ age, *Answer Key*: space, H.M.’s *Response*: stone; *Clue*: change, in a way, *Answer Key*: adapt, H.M.’s *Response*: alter; *Clue*: celestial body, *Answer Key*: moon, H.M.’s *Response*: star; *Clue*: small insect, *Answer Key*: mite; H.M.’s *Response*: flea).

How accurate was H.M.’s comprehension of the crossword clues?

Many different types of clues occur in a crossword puzzle. We were interested in the distribution of H.M.’s errors across the four clue types: Synonyms, Definitions, Categories, or Fill-in-the-Blanks. H.M. responded consistently with valid words for all the different types of clues (Fill-in-the-Blanks = 96%; Categories = 94%; Synonyms = 90%; Definitions = 86%; Overall = 91%) (Figure 1). Additionally, his non-pronounceable nonwords were not confined to any one clue type (Fill-in-the-Blanks = 1%; Categories = 2%; Synonyms = 1%; Definitions = 2%; Overall = 2%), indicating that

*Clue*: dough whippers, *Answer Key*: beaters, H.M.’s *Response*: beeters.)
he did not exhibit disproportionate difficulties with any of these semantic tasks.

When the answers required a proper noun or a proper adjective, H.M. responded accordingly, even when he made errors. This analysis relied on the word itself (because nearly all of his answers were printed in capital letters). We had to decipher misspellings and incorrect proper names in some cases. However, we did not observe common nouns written for clues that required proper ones. Thus, H.M. appreciated the status/role of proper nouns and adjectives.

Different clues also required different parts of speech. We asked whether H.M. showed preserved grammar when he made an error. Did he respond with the same part of speech as that of the answer key? The proportion of responses was also strikingly similar (Figure 2). Of the 732 cases in which he provided an alternative response that was a valid word, he responded with the same part of speech 72% of the time. By chance, the value would be 36%. In the instances when he did not use the same part of speech, many of the clues had multiple interpretations. In some circumstances, H.M.’s choice of a different part of speech was warranted (Table 1). Thus, some of his answers might just as easily have been engendered by the ambiguous nature of the clues rather than by a deficit in his grammar.

**DISCUSSION**

We studied H.M.’s language skills using an ecologically valid approach—that is, on a self-chosen task carried out in his leisure time without pressure. From this, we have found that H.M. did not have any unusual difficulties with the orthographic and grammatical components of language needed for crossword puzzles. Consistent with previous studies, he exhibited few spelling errors (Kensinger et al., 2001). In addition, we found that he responded with a proper noun or proper adjective, when appropriate; and, if his response differed from that given in the answer key, he almost always responded with a valid word in the appropriate part of speech. To this extent, H.M. solves crossword puzzles competently and

**TABLE 1**

Examples in which H.M.’s answer differed from that of the Answer Key in relation to part of speech

<table>
<thead>
<tr>
<th>H.M.</th>
<th>Answer Key</th>
<th>Clue</th>
</tr>
</thead>
<tbody>
<tr>
<td>them</td>
<td>she’s</td>
<td>“_____ not you”</td>
</tr>
<tr>
<td>wild</td>
<td>wadi</td>
<td>African landscape feature</td>
</tr>
<tr>
<td>man</td>
<td>err</td>
<td>be human</td>
</tr>
<tr>
<td>blue</td>
<td>pale</td>
<td>beyond the _____</td>
</tr>
<tr>
<td>alter</td>
<td>dimes</td>
<td>change</td>
</tr>
<tr>
<td>rahs</td>
<td>bars</td>
<td>Cheers and others</td>
</tr>
<tr>
<td>large</td>
<td>titan</td>
<td>giant</td>
</tr>
<tr>
<td>clasp</td>
<td>tress</td>
<td>lock</td>
</tr>
<tr>
<td>delaying</td>
<td>holdover</td>
<td>postpone</td>
</tr>
<tr>
<td>poor</td>
<td>rags</td>
<td>riches’ opposite</td>
</tr>
<tr>
<td>yes</td>
<td>wed</td>
<td>said “I do”</td>
</tr>
<tr>
<td>most</td>
<td>hogs</td>
<td>takes the lions share</td>
</tr>
<tr>
<td>radio</td>
<td>relay</td>
<td>transmit</td>
</tr>
<tr>
<td>real</td>
<td>very</td>
<td>truly</td>
</tr>
<tr>
<td>sail</td>
<td>trip</td>
<td>voyage</td>
</tr>
<tr>
<td>oaths</td>
<td>swear</td>
<td>vow</td>
</tr>
<tr>
<td>alas</td>
<td>ills</td>
<td>woes</td>
</tr>
</tbody>
</table>
demonstrates the necessary word-retrieval skills. This result is consistent with a previous study in which H.M. performed between one and two standard deviations of the means of healthy volunteers on various measures of cognitive function believed to underlie skill in solving crossword puzzles (Skotko et al., 2004).

In the present study, we found that the word frequency of H.M.’s mistakes was higher than those of the answer key, but this is to be expected. Higher-frequency words come to mind more easily (Rubin, 1983), implying that persons without amnesia would also have generated mistakes with a higher frequency score. An additional explanation could be that neural deficits could lead to difficulty with words used less often (MacKay et al., 1998a, 1998b; MacKay & James, 2001). What is interesting about H.M.’s word-frequency scores, however, is the fact that not all of his responses are high-frequency words. If H.M. had an undeveloped vocabulary, we might suspect that all of his mistakes would be words that were more common in the English language than those of the typically obscure crossword answers. However, for every two errors H.M. produced that were more common, one was produced that was of lower word frequency, suggesting that his vocabulary could be, at times, more erudite than that of the key.

Four judges rated H.M.’s responses as more convincing than those of the answer key about 33% of the time. However, H.M. was still challenged by his memory limitations. Several of the clues relied unmistakably on information after 1953 for the correct answers (e.g., Clue: Cheers and others, Answer Key: bars; Clue: _____ age, Answer Key: space; Clue: kind of cone, Answer Key: nose). In these instances, H.M. either chose not to respond or creatively generated a feasible answer using his preoperative memory (e.g., in the case of the above, H.M.’s responses were “rahs”, “stone”, and “pine”, respectively). However, these examples further support the claim that H.M.’s lexical word-retrieval skills remain fluid despite his profound anterograde amnesia.

In those infrequent instances in which he responded with a pronounceable or non-pronounceable nonword, H.M. often appeared to have generated a clever truncation or alteration of a valid word to fit the space constraints of the puzzle. From his behaviour in the laboratory and our observation of his completed puzzle books, it appears that H.M. solved puzzles in a linear fashion—completing the Across clues before proceeding to the Down clues—and, by doing so, he did not benefit from crisscrossing letters when completing the Across clues. Such a maladaptive strategy could be a result of H.M.’s personal preference, his memory limitations in going back and forth between Across and Down clues, inflexibility following from frontal-executive dysfunction (Hebben, Corkin, Eichenbaum, & Shedlack, 1985; Kensinger et al., 2001), or some combination of the three possibilities. Some of H.M.’s non-pronounceable nonwords might also have been residues of answers that went unchecked. For example, by completing a series of Across clues, he might have also completed a Down clue. If he neglected to check this Down clue, the letter strings from the completed Across clues may have formed a non-pronounceable nonword as an answer to a Down clue. However, he rarely generated these types of mistakes.

In sum, H.M.’s orthographic and grammatical use on his recreational crossword puzzles supports the view that he has a strong command of comprehension and lexical access of the English language. Our findings contribute more evidence to the question of H.M.’s language skills, not addressed by previous studies, by offering important evidence on language comprehension, word-retrieval skills, lexical access, and grammar usage. This study is limited by the single-subject design. However, given the nature of H.M.’s amnesia, his inestimable contribution to our understanding of memory, and the current debate about his language skills, we believe that in spite of this limitation, our examination of his crossword puzzles is revealing. Previous studies on H.M.’s language have already compared him with age- and education-matched healthy volunteers (Kensinger et al., 2001; Skotko et al., 2005).

The current study and previous ones provide a fuller picture of H.M.’s language skills and show the limited role that the MTL could play in the maintenance of language skills over the adult lifespan. For H.M., this is a good thing, as he continues to complete books of crosswords with a level of accuracy that bring him satisfaction.
REFERENCES


