No Sex Differences in Memory Formation during General Anesthesia


Background: Women respond differently to anesthesia than men, initially recovering more rapidly, but having more postoperative morbidity. Studies on surgical patients report evidence of memory formation during anesthesia. However, sex differences in memory formation have not been explored. Therefore, the authors investigated sex differences in the implicit and explicit memory formation during general anesthesia.

Methods: With ethics committee approval, 120 consenting adult patients scheduled to undergo surgery during general anesthesia were recruited. Intraoperatively, 16 target words were presented to patients via headphones, and the Bispectral Index was recorded. Postoperatively, memory for presented words was tested using a word stem completion test. The test was divided into inclusion and exclusion parts, to delineate implicit and explicit memory contributions.

Results: Target and distracter hit rates were similar in men and women. For the whole study group, there was a significant difference between inclusion target hit rate (0.42) and base hit rate (0.59) (P = 0.01). Buchner's model suggested that this memory formation was attributable to both implicit and explicit memory. A Bispectral Index value greater than 50 was the only significant predictor of inclusion target hit rate. None of the patients were able to consciously recall the words presented during surgery.

Conclusions: Patients showed greater memory performance for words presented during general anesthesia than for words not presented. However, sex differences in memory formation were not observed. A relation between hypnotic state and memory during sevoflurane anesthesia was also established, suggesting that memory formation is possible even at hypnotic depths considered to be adequate anesthesia.

WHETHER patients may form unconscious memories during anesthesia, and whether these memories matter, is an intriguing subject.1,2 Older studies on this subject were confounded by methodologic problems, including inability to correlate memory with depth of anesthesia because of inadequate depth of anesthesia monitoring, and inability to eliminate the influence of active recall when testing for unconscious memory.3,4 Recent studies have used Bispectral Index (BIS) monitoring and a sophisticated memory-testing technique called the Process Dissociation Procedure5 to overcome these problems.

These studies report that memories can be formed unconsciously during apparently adequate anesthesia, but only if anesthesia is relatively light6–8,11,15 and word presentation occurs during surgery.6–12,13 However, these factors explain only part of the variability between patients with respect to unconscious memory formation, and more predictive factors must be sought.1,2

Women recover more rapidly from anesthesia than men.16–24 This phenomenon has not been fully explained but may mean that women regain cognitive function more rapidly than men if anesthetic depth lightens during surgery. Indeed, women are more likely to report awareness25–29 and dreaming30–32 during anesthesia than men, but whether women form more unconscious memories during anesthesia than men is not known. We therefore tested the hypothesis that women have a greater hit rate than men for target words in a word stem completion (WSC) test administered during a standardized BIS-titrated anesthetic. We also explored the contribution of explicit and implicit memory to general memory performance using the Process Dissociation Paradigm and determined the relation between memory formation and the depth of anesthesia as determined by BIS in our patients.

Materials and Methods

With the approval of the Human Research Ethics Committee of the Royal Melbourne Hospital (Parkville, Victoria, Australia) and written informed consent, 120 patients aged 18–75 yr, with American Society of Anesthesiologists’ physical status I–III, and presenting for elective surgery during relaxant general anesthesia were studied. All participants were native English speakers or fluent in English as their second language. Exclusion criteria included (1) intracranial, aural, and cardiothoracic surgery; (2) deafness; (3) memory disturbance or intellectual disability; (4) plan for benzodiazepine premedication; and (5) not expected to be available for, nor able to cooperate with, the postoperative interview.

Word Stem Completion Test

We used the WSC test (a commonly used perceptual priming task6,7,11–15) to measure memory performance. In the WSC test, the patient is primed by exposure to a target word during anesthesia. After recovery from anesthesia, the patient is presented with the first part of a word (word stem) and is asked to complete it. Comple-
tion of a word stem with a target word presented during surgery is termed a “hit.” An increased hit rate over the base rate without priming is evidence of memory formation.

Thirty-two five-letter words with a base hit rate of approximately 0.30 were selected from a list compiled by us in Australian patients. Each target word was recorded digitally by the same female investigator and then edited with Cool Edit 2000 1.1 (Syntrillium Software Corporation, Phoenix, AZ), so that only the first three letters were enounced and a word stem was derived. Audio compact disks were created containing a short greeting and complete words, repeated 40 times each for presentation during anesthesia, and the appropriate word stems for presentation postoperatively.

Based on a previous study by Lubke et al., who demonstrated learning (hits = 6.9 words out of 16) over base rates (hits = 4.8 words out of 16) in women undergoing cesarean delivery during general anesthesia, we planned to recruit 60 male and 60 female patients (α = 0.05; β = 0.2; SD = 2.0). Patients underwent block randomization by sex to one of four groups (n = 30 per group). Each group contained 15 men and 15 women. Two lists of target words and two lists of distracter words were allocated to each group (table 1). Ten compact discs were created for each group with target words and word stems randomized in order on each compact disc, and these were allocated randomly to each patient in the group. Each compact disc was used three times. The word presentation lasted approximately 42 min. Patients and postoperative observers were blind to the allocation of the word lists.

### Procedure

Before induction of anesthesia, patients were asked to adjust headphone volume and position for maximum comfort. Routine monitoring and BIS monitoring (BIS-XP® version 4.0; Aspect Medical Systems, Inc., Newton, MA) were commenced. Anesthesia was induced with fentanyl, propofol, and a nondepolarizing muscle relaxant and maintained with sevoflurane in oxygen-air. The trachea was intubated, and intermittent positive-pressure ventilation was commenced via a circle circuit. Anesthesiologists were asked to titrate the inspired concentration of sevoflurane in a BIS range of 55–60 during word presentation to ensure that subjects were standardized to the depth of hypnosis. Morphine and antiemetics were administered after word presentation, if indicated. After the first skin incision, the two lists of words were played to the patient via a laptop computer that also recorded BIS values at 5-s intervals. At the end of surgery, neuromuscular blockade was reversed, the patient’s trachea was extubated, and he or she was transferred to the postanesthesia care unit.

Patients were assessed 4–24 h postoperatively when they recovered from the hypnotic effects of anesthesia (i.e., were awake and orientated with pain and nausea adequately controlled). The WSC test was then administered. Instructions were provided before each part of the test. The word stems from all four lists were presented aurally via headphones, maintaining the same context of presentation as during surgery. Patients were asked to write down their responses, or if necessary, the researcher helped to record their responses. We used the Process Dissociation Paradigm to separate implicit and explicit learning. In the inclusion part, patients were asked to complete word stems with a word that they remembered hearing during anesthesia or the first word that came to mind (one list of target words and one list of distractors). In the exclusion part, patients were asked to complete word stems with a word they had not heard during anesthesia (one list of target words and one list of distractors). Implicit memory facilitates a higher hit rate in the inclusion and exclusion parts over the base rate, whereas explicit memory increases the inclusion hit rate but decreases the exclusion hit rate.

The following data were collected: demographics, preoperative quality-of-recovery scores (a validated nine-item questionnaire on quality of recovery with a minimum score of 0 and a maximum score of 18; collection of a preprocedure baseline score is also validated) and hospital anxiety and depression scores (0–14 possible score for each of anxiety and depression), times of induction of anesthesia, first surgical incision, completion of wound closure, tracheal extubation, first eye opening, and Aldrete score of greater than 9. Before postoperative memory testing, patients were assessed for quality-of-recovery scores, hospital anxiety and depression scores, recall of anesthesia, and dreaming using a standardized questionnaire: “What is the last thing they were presented with lists 1 and 3 during surgery. They were tested on lists 3 and 4 during the inclusion part and lists 1 and 2 during the exclusion part of the test.

### Table 1. Word List Presentation Scheme

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>Inclusion target</td>
<td>Inclusion distracter</td>
<td>Exclusion target</td>
<td>Exclusion distracter</td>
</tr>
<tr>
<td>Group 2</td>
<td>Exclusion distracter</td>
<td>Inclusion target</td>
<td>Inclusion distracter</td>
<td>Exclusion target</td>
</tr>
<tr>
<td>Group 3</td>
<td>Exclusion target</td>
<td>Inclusion distracter</td>
<td>Exclusion target</td>
<td>Inclusion distracter</td>
</tr>
<tr>
<td>Group 4</td>
<td>Inclusion distracter</td>
<td>Exclusion target</td>
<td>Exclusion distracter</td>
<td>Inclusion target</td>
</tr>
</tbody>
</table>

Word presentation counterbalancing scheme (as per Lubke et al.8). Each patient was randomly assigned to a group. During surgery, the two lists of target words were presented. Postoperatively, all four lists (target and distracters) were used for the word stem completion test. For example, if a patient was assigned to group 3, he or she was presented with lists 1 and 3 during surgery. They were tested on lists 3 and 4 during the inclusion part and lists 1 and 2 during the exclusion part of the test.
you remember before you went to sleep?” “What is the first thing you remember when you woke up?” “Did you remember anything in between?” and “Did you have any dreams during anesthesia?”

**Statistical Analyses**

Statistical analyses were performed using Stata 8.2 (Stata Corporation, College Station, TX). Continuous variables were graphed to assess their distribution. Normally distributed variables were summarized using mean ± SD, skewed variables using median (range), and categorical variables using number (%). Comparisons were made using the Student t test (normally distributed variables), the Wilcoxon rank sum test (skewed variables), or chi-square tests (categorical variables). Survival data (times to an event) were assessed using log-rank tests.

To assess the constancy of BIS values, we calculated the performance error (PE: bias), using the formula PE = ((BIS - mean BIS)/mean BIS) × 100, the absolute PE (inaccuracy), which is the absolute value of the PE and the wobble (intraindividual variability) using the formula wobble = |PE − median PE|.

The mean numbers of hits, and the hit probability for word stems that were attempted, were calculated for targets and distracters in the inclusion and exclusion parts of the test. The comparisons of interest were (1) inclusion target hit rate versus inclusion distracter hit rate (testing general memory performance) and (2) exclusion target hit rate versus exclusion distracter hit rate (testing explicit memory performance). These comparisons were made using paired, two-tailed t tests. Figure 1 shows the Buchner multinomial processing model, which was used to estimate explicit and implicit memory probabilities from the hit frequencies for targets and distracters in the inclusion and exclusion tests.

Bispectral Index values were categorized (≤ 40, 40.1–45, 45.1–50, 50.1–55, 55.1–60, > 60) on the basis that the range 40–60 is considered appropriate for general anesthesia. By definition, distracters were not presented during surgery and therefore had no corresponding BIS value, so for all BIS categories, matching distracter frequencies were calculated from base rate performance (i.e., 0.38) and the number of words presented. We addressed whether general memory performance was significant at BIS values greater than or less than 50, on the basis that (1) it represents the midpoint of the recommended range; (2) learning seems more likely with BIS values above 50; and (3) there have been previous reports of wakefulness and awareness in the 50–60 range or below. Chi-square analyses were used to compare hits/misses in the target/distracter groups at BIS values greater than or less than 50.

Univariate logistic regression was used to determine predictors of general memory performance (as determined by the hit rate on the inclusion target words that were attempted). Prospectively defined predictors included age, sex, preoperative hospital anxiety and depression score, mean BIS value greater than 50 during word presentation, duration of anesthesia, time to eye opening, time to WSC test, and report of dreaming. Continuous variables were categorized based on visual inspection of graphs versus hit rate. Predictors with P values less than 0.2 were included in a multivariate logistic regression model.

**Results**

Recruitment to the trial is outlined in figure 2. Six men and one woman received midazolam before induction despite its exclusion from the protocol and were removed from the analyses. The 52 men and 61 women remaining were aged 48 ± 15 and 45 ± 15 yr (P = 0.32), weighed 86 ± 15 and 76 ± 18 kg (P = 0.002), and had mainly American Society of Anesthesiologists physical status I or II (79% and 82%; P = 0.96), respectively. Mean BIS values were similar in men and women during word presentation but were lower than suggested by the protocol (table 2). The median of the median PEs during word presentation was 0% (range, −18% to 8%), and the median of the median absolute PEs was 8% (1% to 24%). The median of the median wobbles was 7% (0% to 24%), i.e., the typical patient’s BIS value varied by 3 BIS units during the word presentation, and the worst patient’s BIS varied by 14 BIS units.

Follow-up interviews were conducted on men and women 16.1 (1.5–24) and 17.4 (3.8–24) hours postop-
eratively ($P = 0.23$). Postoperative quality-of-recovery scores were significantly lower than preoperative scores in men ($P = 0.003$) and women ($P < 0.001$). No significant differences in hospital anxiety and depression scores were observed in either men or women ($P = 0.53$ and $P = 0.43$, respectively). The prevalence of dreaming among interviewed patients was $11\%$ (men $10\%$ and women $12\%$; $P = 0.68$). None of the patients were able to consciously recall the words presented during surgery. One woman undergoing a laparoscopic cholecystectomy recalled a short episode involving a tugging sensation in the right upper quadrant of her abdomen and being unable to move. She did not recall hearing the word presentation. The recorded BIS data showed several 1- to 2-min periods of BIS values between 60 and 63. She was interviewed on several further occasions and was offered counseling, and there were no adverse sequelae.

The hit rates and frequencies were calculated based on the number of word stem completion attempts, because 119 presented stems (4%) were not completed. No differences were observed in hit rates between men and women in either of the four test conditions. Target words were used to complete word stems significantly more often than distracter words in the inclusion part of the test ($P = 0.01$), but there was no significant difference between target and distracter hit rates in the exclusion part of the test ($P = 0.30$) (table 3). Although the observed base hit rate in the inclusion test (0.39) was higher and that in the exclusion test (0.24) was lower than 0.3 (the base rate as evidenced in our pilot study, which was conducted in a demographically similar cohort of patients), these differences were not significant ($P = 0.30$ and $P = 0.29$, respectively).

The estimate for explicit memory from the Buchner model was 0.038 (SE = 0.018 [95% confidence interval, 0.002–0.073]) and for implicit memory was 0.041 (SE = 0.019 [0.004–0.078]). The 95% confidence interval for both explicit and implicit memory estimations were exclusive of zero (i.e., significant).

General memory performance was significantly higher than base rate performance at BIS values greater than 50
At BIS values of less than 50, general memory performance was not significantly different from base rate performance ($P = 0.83$). A BIS value greater than 50 was the only predictor of inclusion target hits (i.e., improved general memory performance) in the logistic regression model (tables 4 and 5 and fig. 3).

### Discussion

This study provides further evidence that memory formation may occur during apparently adequate general anesthesia.6–8,10–14 However, we were unable to demonstrate a difference in memory formation between men and women. We conclude that men and women anesthetized to equivalent levels of hypnosis with sevoflurane are equally likely to form memories during anesthesia.

Target hit rates were higher than distracter hit rates in the inclusion part of our study, indicating improved general memory performance over baseline. In contrast, target and distracter hit rates in the exclusion part were similar, providing no evidence for explicit memory formation. However, exclusion target and distracter hit rates were lower than expected from the pilot study.33 This may be because our patients completed fewer exclusion stems than inclusion stems and tended to use unusual words to complete the exclusion part of the test. By including data from both parts of the test in the Buchner model, we were able to differentiate the effects of implicit and explicit memory on general memory performance and detect both implicit memory and a weak form of explicit memory that allowed exclusion decisions to be made.

In addition, we were able to explore the relation between memory formation and anesthetic depth by matching BIS values to each word presentation. Our results suggest that the odds of memory formation are significantly increased at BIS values greater than 50. This is consistent with previous reports,6–8,11,14 in particular the study of Lubke et al.7 in cesarean delivery patients. In that study, inclusion hit rates were significantly greater than, whereas exclusion hit rates were not different from, base rates, and the Buchner model revealed a form of explicit memory that allowed inclusion–exclusion decisions to be made in the absence of postoperative recall. The mean BIS value during word presentation was 76. Similarly, Deeprose et al.12,13 reported implicit memory formation during surgery even when BIS values were lower (mean = 42). In contrast, Kerssens et al.15 were unable to demonstrate memory formation when mean BIS values during word presentation were 49 (identical to our study). The rate of fluctuation in anesthetic depth during word presentation may be a crucial factor: 8% of words were presented as BIS values greater than 60 in our study, 11% in the study of Deeprose et al.,13 but only 1% in the study of Kerssens et al.15 Greater fluctuation in our study compared with that of Kerssens et al. may have been due to our higher target BIS range of 55–60.

One woman in our study had an episode of awareness.

### Table 3. Word Stem Hit Rates

<table>
<thead>
<tr>
<th></th>
<th>Inclusion Targets</th>
<th>Inclusion Distracters</th>
<th>Exclusion Targets</th>
<th>Exclusion Distracters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Men (n = 47)</td>
<td>Women (n = 56)</td>
<td>Men (n = 47)</td>
<td>Women (n = 56)</td>
</tr>
<tr>
<td>Hit rate</td>
<td>0.42 (0.18)</td>
<td>0.45 (0.16)</td>
<td>0.39 (0.21)</td>
<td>0.27 (0.20)</td>
</tr>
<tr>
<td>Hit rate*</td>
<td>0.44 (0.17)</td>
<td>0.39 (0.19)</td>
<td>0.26 (0.18)</td>
<td>0.24 (0.18)</td>
</tr>
<tr>
<td>Hit frequency</td>
<td>152</td>
<td>197</td>
<td>138</td>
<td>176</td>
</tr>
<tr>
<td>Hit frequency*</td>
<td>352</td>
<td>311</td>
<td>204</td>
<td>190</td>
</tr>
<tr>
<td>$P$ value†</td>
<td>0.38</td>
<td>0.86</td>
<td>0.52</td>
<td>0.26</td>
</tr>
</tbody>
</table>

* Men and women combined. † Between men and women within each part of the test.

$P = 0.012$). At BIS values of less than 50, general memory performance was not significantly different from base rate performance ($P = 0.83$). A BIS value greater than 50 was the only predictor of inclusion target hits (i.e., improved general memory performance) in the logistic regression model (tables 4 and 5 and fig. 3).

### Table 4. Predictors of General Memory Performance

<table>
<thead>
<tr>
<th>Predictor Variable</th>
<th>Odds Ratio (95% Confidence Interval)</th>
<th>$P$ Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, yr</td>
<td>0.99 (0.98–1.00)</td>
<td>0.22</td>
</tr>
<tr>
<td>Female sex</td>
<td>1.18 (0.90–1.56)</td>
<td>0.39</td>
</tr>
<tr>
<td>Preoperative HAD score</td>
<td>0.99 (0.96–1.01)</td>
<td>0.33</td>
</tr>
<tr>
<td>Bispectral Index &gt; 50</td>
<td>1.37 (1.04–1.80)</td>
<td>0.02</td>
</tr>
<tr>
<td>Duration of anesthesia, min</td>
<td>1.00 (0.99–1.00)</td>
<td>0.42</td>
</tr>
<tr>
<td>Time to eye opening, min</td>
<td>0.98 (0.96–1.00)</td>
<td>0.38</td>
</tr>
<tr>
<td>Time to word stem completion test, min</td>
<td>1.00 (0.99–1.00)</td>
<td>0.94</td>
</tr>
<tr>
<td>Dreaming, yes</td>
<td>0.73 (0.47–1.13)</td>
<td>0.27</td>
</tr>
</tbody>
</table>

HAD = Hospital Anxiety and Depression. A high score indicates more anxiety and depression.
Because she did not remember word presentation, it is possible that this episode occurred during a 10-min period toward the end of surgery when BIS values were around 60. No other signs of inadequate anesthesia were detected during the operation. This episode adds weight to an argument that episodes of BIS greater than 60 must be interpreted promptly along with other data and be acted upon if necessary. Furthermore, we suggest that BIS values less than 50 (rather than less than 60) may be necessary to make awareness and unconscious memory formation a sufficiently remote possibility for our patients.

This study was designed to maximize our ability to demonstrate a difference between memory formation during anesthesia in men and women, if one existed: (1) We avoided the use of preoperative midazolam, because of its amnesic and anxiolytic effects; (2) we chose sevoflurane (rather than propofol) for anesthetic maintenance; (3) we standardized anesthetic depth at 55–60; (4) we presented words during (rather than before) surgery; (5) we attempted to test patients as soon as they recovered from the hypnotic effects of anesthesia; (6) we maintained the same (auditory) context during priming and testing; and (7) we used the Process Dissociation Paradigm to distinguish between the contributions of implicit and explicit memory.

Several issues arise from our study design that require discussion. Firstly, six patients received midazolam despite its exclusion from the protocol. However, post hoc analyses excluded an influence of midazolam on the primary endpoint. Secondly, a BIS target of 55–60, although likely to maximize memory formation in this context, was not achieved in practice. Anesthesiologists encountered difficulty in titrating sevoflurane to maintain the BIS in this range, in the face of varying surgical stimulation, the lag between alterations in dialed concentration and effect, the lag in calculating the BIS, and concern about increased excursions of BIS values above 60. The lower mean BIS values achieved in this study (men, 48; women, 49) may have minimized the difference between men and women in memory formation, because fewer patients may have had arousals during anesthesia that allowed memories to be formed.

Thirdly, we chose to interview patients between 4 and 24 h postoperatively. Because the effects of priming fade with time, this may have minimized our estimate of learning, but not the difference between men and women. Although a response to priming may still be demonstrated after 48 h, as short a period as possible between priming and testing is preferable, although testing while residual hypnotic effects of anesthesia remain is undesirable.

Recovery times and the quality of recovery, as evidenced by quality of recovery and hospital anxiety and depression scores, were similar in men and women. Our results contrast with previous reports of more rapid awakening from anesthesia maintained with propofol or volatile anesthetics, but poorer quality of recovery, in women than in men. Recovery from anesthesia was defined as the time from cessation of sevoflurane to the time of first eye opening. However, administration of sevoflurane after the word presentation was completed was not done per protocol, so some patients may have received reducing concentrations of sevoflurane before cessation, whereas others may have been on constant concentrations. The amount of opioid analgesia also was not controlled. The pattern of drug administration could have been different in women and men, and this could have influenced this result. Alternatively, this study may have been underpowered to assess this secondary endpoint.

In conclusion, we demonstrated improved general memory performance after priming during general anesthesia with sevoflurane, but there was no difference in intrinsic memory formation between men and women.
between men and women. Memory formation was more likely at BIS values greater than 50. The clinical importance of implicit memory formation during anesthesia remains unknown, and further research is required to establish whether it has any untoward sequelae.1,2

References

1. Veselis R: Gone but not forgotten—or was it? Br J Anaesth 2004; 92:161–3