Category generation testing in the search for implicit memory during general anaesthesia

W. J. MACRAE, J. M. THORP AND K. MILLAR

Summary
We gave auditory examples of two semantic categories through headphones to 100 surgical patients anaesthetized with propofol and enflurane. This presentation was made during certain stages of the procedure, potentially associated with arousal, and during steady-state anaesthesia. Postoperative review using category generation tests showed successful priming in a pre-induction group but no evidence of implicit memory in the anaesthetized groups. These results suggest that timing an auditory input to coincide with surgical stimulation does not increase the probability of retrieval of information by this type of testing. (Br. J. Anaesth. 1998; 80: 588–593)

Keywords: anaesthetics volatile, enflurane; memory, implicit; anaesthesia, depth

Intraoperative awareness occurs if the intensity of the surgical stimulation is not matched by an appropriate level of anaesthesia. Alternatively, faulty anaesthetic technique or equipment may be responsible for an alert state and subsequent memory for the event. The sensitivity of testing of recall has been increased by the division of memory into explicit and implicit components.1 Explicit memory is conscious recall of information and often of the events surrounding the original learning. In contrast, implicit memory involves the passive and unconscious activation of existing memory traces but without subsequent conscious memory of the initial activating incident. This process is also known as priming because it makes subsequent spontaneous recall of the activated memory more likely. The occurrence of implicit memory under surgical anaesthesia might imply an undesirable lightening of anaesthesia or, alternatively, it might be of therapeutic value.2 3 In most studies, the terms implicit learning and implicit memory are conflated. If learning of auditory information occurs during a period of wakefulness of which the patient is subsequently amnesic, explicit learning with subsequent implicit memory may be said to have occurred. On the other hand, if learning of information has occurred while unconscious, both learning and memory were implicit.

Studies of implicit memory under anaesthesia have yielded both positive1–8 and negative9–10 results. The inconsistency of findings may be because the majority of studies have presented the information to patients throughout the surgical procedure. For those studies with positive results, it is not possible to establish whether priming was most effective at a given stage in the anaesthetic procedure. (A recent review of memory and learning11 highlighted the need for further study to establish the salience of stimulating events in this respect.) It has been suggested12 that there may be specific times during a surgical procedure when potential wakefulness can occur. If wakefulness is a prerequisite for the formation of implicit memory, it would be logical to examine the results of auditory stimuli presented at those particular times for evidence of differential effects.

An early study by Eisele, Weinreich and Bartle13 attempted to show that recall after operation could be related to the time during operation when information was presented. In their study, patients were allocated randomly to one of five groups. The auditory stimulus was presented to each group at one of the following times: before induction, during operation, before extubation, in the recovery room or when fully awake after operation. Some patients, only from the intraoperative group, were monitored for galvanic skin response (GSR) as an indicator of arousal. In these patients, the stimulus was presented only when GSR indicated arousal. No evidence of intraoperative recall was shown despite GSR-indicated arousal. The groups who had heard the stimulus before operation or when fully recovered had complete recall, while the pre-extubation and recovery room groups had impaired recall. This particular study was criticized14 15 for assessing only depth of anaesthesia in a minority of patients when information was being presented. Moreover, the sensitivity of memory tests has subsequently improved.

We have used category generation testing as a test of implicit memory, which has been shown previously to produce positive results during anaesthesia.5 6 Also, we investigated if the occurrence of implicit memory varied as a function of the time during the procedure when the auditory information was presented. Steady-state anaesthesia was also compared with times of potential wakefulness, such as intubation, incision and recovery, when implicit memory might be established.

W. J. MACRAE*, MB, CHB, FRCA, J. M. THORP, MB, CHB, MRCP, FRCA, Monklands and Bellshill Hospitals NHS Trust, Monkscourt Avenue, Airdrie, Lanarkshire ML6 0JS. K. MILLAR, BA, PhD, C. PSYCHOL, FBPSS, Department of Psychological Medicine, University Academic Centre, Gartnavel Hospital, 1055 Great Western Road, Glasgow G12 0XH. Accepted for publication: December 22, 1997.

*Present address for correspondence: Stobhill Hospital NHS Trust, 133 Balornock Road, Glasgow G21 3UW.

This article is accompanied by Editorial I.
Patients and methods

We studied surgical patients undergoing elective or emergency procedures during general anaesthesia. The study was approved by the area Ethics Committee and all patients gave informed consent. Patients were aged 16–70 yr, ASA I–III. Exclusion criteria were hearing difficulties, history of psychotropic or benzodiazepine drug use, local anaesthetic use before skin incision and if English was not the first language. As enflurane was used for anaesthesia, patients with renal failure (acute or chronic) or a history of epilepsy were also excluded. Computer-generated random numbers allowed allocation to one of five groups for the timing of the auditory stimulus.

The category priming and generation task was adapted from Jelicic, Bonke and Appleboom. The category generation task requires the subject to generate a list of exemplars of a named category as they come spontaneously to mind. When collated across a group of individuals, such lists often show remarkable similarity in the first three or four words generated (e.g. for “animals”, most people immediately generate dog, cat, horse, sheep, etc, as being familiar exemplars). Other exemplars (e.g. tiger, camel), while still familiar, are less likely to be produced early in a generation sequence, if at all.

Prior exposure to category exemplars often results in a greater probability of their being generated in a generation task, and to their being among the first words and in a sequence, even when the subject has no conscious recall of prior exposure. Clearly, if priming is to be demonstrated, it is pointless to choose highly familiar target words which have a high probability of being generated even without priming. Relatively less familiar target words should be chosen so that if priming is effective, there is the potential for them to be generated more often and earlier in the generation sequence than would otherwise be the case.

A preliminary study in a similar surgical population of 50 patients allowed selection of three familiar examples of each of two semantic categories “fruit and colours”. Examples of these categories were selected for the population in this area and were designated as target words. The words were selected by the investigators and had been produced by >50% of patients but did not commonly occur as one of the first three words generated (i.e. in serial positions one, two or three). The chosen examples of the two categories, fruit (peach, grape and melon) and colours (black, pink and purple), were recorded onto tape for use in the main study. This information was also used subsequently in the control group for comparison with the experimental groups. A further category “vegetables” was also generated, and these were used as control words in the postoperative testing.

The information was presented via a personal stereo set with headphones. The tape was recorded professionally with a male voice and the first part of the tape was an introduction for the investigator to check satisfactory function before placing the headphones on the patients for the 5-min recording. Each sequence of target words was preceded by the phrase “Please listen carefully to these words...”. The control category, vegetables (onion, broccoli and cauliflower) was not presented to the patients during operation.

Patients were premedicated with pethidine and Phenergan (Pamergan) i.m., 60–90 min before operation. A standard anaesthetic consisted of an induction dose of propofol and maintenance with enflurane and 50–60% nitrous oxide in oxygen, with morphine for analgesia; neuromuscular block was not an exclusion criterion. The aim was to study adequate anaesthesia and this was based on clinical requirements rather than on a weight basis. The tape was played to each patient at one of the following times only: (1) pre-induction, (2) post-induction: immediately after intubation–placement of the laryngeal mask, (3) post-induction: immediately after the first incision, (4) steady state: 10 min after the first incision, and (5) recovery: enflurane administration discontinued and after breathing 100% oxygen for 2 min, with an end-tidal enflurane concentration of 0.1–0.3%. End-tidal enflurane concentrations were noted at the beginning and end of tape playing, after which the headphones were removed. The actual times of starting the cassette tape, cessation of anaesthesia and postoperative testing were recorded.

Category generation testing was carried out 5–24 h after the end of anaesthesia. Excluding 8 h for sleep, all patients were tested within 14 h of the auditory stimulus. The order in which patients free-associated to each category was randomized. Patients were asked to name examples (up to 10) of each category as they came to mind (including the control category which had not been presented during anaesthesia). The examples were written down by the interviewer and 2 min were allowed for each category. No mention was made of the taped information at this stage. Patients were subsequently asked to recall being in the anaesthetic room, their next memory after this (to assess explicit memory) and whether they had any recollection of information from the tape played during anaesthesia.

Data from category generation were analysed as follows. The hits score was the total number of target words generated. The serial position score took account of the position of a target in the list of generated words. The sequence rate defined those occasions where two or three targets were generated one after another. The null hypothesis was that there would be no difference either between the control and experimental groups or between the control category and target categories. Results were analysed with Minitab release 7.2, using two-tailed, non-parametric statistics, with adjustment for ties, where appropriate. Given the potential for multiple comparisons leading to type 1 errors, \( P < 0.01 \) was set as the criterion for significance. To avoid type 2 errors, \( P \) values of 0.05–0.01 are reported.

Results

We studied 104 patients, of whom 100 completed the study; two patients refused to participate in the study and data were incomplete for two patients. There were no significant differences between groups in age, sex or period of time to testing. Comparisons were made of hits, serial positions and sequences of target word generation by the experimental groups and by the control group, and also between target and control words generated by the experimental groups.
HITS

There was a significant difference in hits between the groups (Kruskal–Wallis, \(P=0.001\)). The pre-induction group generated significantly more hits (table 1) than either the control group or any of the anaesthetized groups (Mann–Whitney, \(P<0.002\) in each case), except for the post-incision group (\(P=0.03\)). When the pre-induction group was omitted from the former analysis, no significant inter-group difference remained. Within the anaesthetized groups, the post-incision group had the highest, and the recovery group the lowest, hit rate, but neither was significantly different from the control group. Comparing the groups for the control category (vegetables), there were no significant differences between groups (Kruskal–Wallis, \(P>0.7\)). The pre-induction group also showed a significantly higher hit rate for the target words than for the control words (Mann–Whitney, \(P=0.0009\)). No difference in hits between target and control words was found for any other group.

SERIAL POSITIONS

There was an overall difference between groups (Kruskal–Wallis, \(P<0.001\)) whereby the serial position of target words generated by the pre-induction group differed significantly from each of the experimental groups and from the control group (Mann–Whitney, \(P\) values all \(<0.0001\)) (table 1). Exclusion of the pre-induction group showed no significant difference between the remaining groups. For the pre-induction group, comparison of serial positions of target words and control words showed that although the target words were generated earlier, this was not significant (Mann–Whitney, \(P=0.015\)). It was interesting that the recovery group, which had the lowest hit rate for the target words, also showed a lower median position for generated target words than for the control words, nearly achieving the set level of significance (Mann–Whitney, \(P=0.012\)).

It was postulated that if implicit memory were occurring, but only in a few patients, then a high hit rate might be associated with a higher median serial position. There was no evidence that this occurred; patients with five or six hits out of the possible six target words did not show any significant difference in serial position compared with similar high-scoring subjects in the control group, nor did they differ from the remainder of the experimental patients with fewer than five or six hits.

SEQUENCES

The pre-induction group generated significantly more sequences of target words than either the control group or any of the other experimental groups (Mann–Whitney, \(P<0.0001\)). There were no differences between the other groups (Kruskal–Wallis). The pre-induction group also showed significantly more sequences of target words than control words (Mann–Whitney, \(P<0.0001\)). For the control group and each of the anaesthetized groups, there was no difference between target words and control words (\(P>0.1\)).

It may be pertinent to view the sequences generated where sequence was a possibility; instances of one or no hit preclude this. In the pre-induction group, only one patient failed to generate at least two of the target words for colours while seven patients failed in this respect in the fruit category, but the

---

**Table 1** End-expiratory enflurane concentration (mean (SD)) during auditory stimulus and subsequent category generation, and serial position (median (interquartile range)) of successfully generated target words and control words

<table>
<thead>
<tr>
<th></th>
<th>Enflurane (%)</th>
<th>Six target words Hits (%)</th>
<th>Serial position</th>
<th>Three control words Hits (%)</th>
<th>Serial position</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control subjects</td>
<td>0.0 (0.0)</td>
<td>64</td>
<td>6 (4–8)</td>
<td>51</td>
<td>6 (4–8)</td>
</tr>
<tr>
<td>Experimental groups</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-induction</td>
<td>0.0 (0.0)</td>
<td>84</td>
<td>4 (2–6)</td>
<td>53</td>
<td>5 (3–7)</td>
</tr>
<tr>
<td>Post-induction</td>
<td>0.34 (0.3)</td>
<td>61</td>
<td>7 (5–8)</td>
<td>56</td>
<td>6 (4–7)</td>
</tr>
<tr>
<td>Post-incision</td>
<td>1.05 (0.2)</td>
<td>70</td>
<td>6 (5–8)</td>
<td>58</td>
<td>6 (4–8)</td>
</tr>
<tr>
<td>Steady state</td>
<td>0.92 (0.2)</td>
<td>60</td>
<td>6 (4–7.7)</td>
<td>48</td>
<td>6 (4–8)</td>
</tr>
<tr>
<td>Recovery</td>
<td>0.28 (0.05)</td>
<td>56</td>
<td>6 (5–8)</td>
<td>63</td>
<td>5 (4–7)</td>
</tr>
</tbody>
</table>

---

**Table 2** Subjects giving two or three words in a sequence as a percentage of \(n\); \(n\) = number of subjects in that group with two or three hits; \(N\) = total number of subjects in that group

<table>
<thead>
<tr>
<th></th>
<th>Colours</th>
<th>Fruit</th>
<th>Control words</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(n/N)</td>
<td>(2) (%)</td>
<td>(3) (%)</td>
</tr>
<tr>
<td>Control subjects</td>
<td>42/50</td>
<td>35</td>
<td>5</td>
</tr>
<tr>
<td>Experimental groups</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-induction</td>
<td>18/19</td>
<td>50</td>
<td>39</td>
</tr>
<tr>
<td>Post-induction</td>
<td>16/21</td>
<td>50</td>
<td>12</td>
</tr>
<tr>
<td>Post-incision</td>
<td>19/20</td>
<td>26</td>
<td>16</td>
</tr>
<tr>
<td>Steady state</td>
<td>16/21</td>
<td>44</td>
<td>12</td>
</tr>
<tr>
<td>Recovery</td>
<td>15/19</td>
<td>0</td>
<td>20</td>
</tr>
</tbody>
</table>
sequence rate was very high in the remaining majority (table 2). The recovery group, which had fewer hits and lower serial positions for target words, also had the fewest sequences of target words. As the hit rate for this group was lower, this might be anticipated. However, excluding instances of none or one hit, this group still demonstrated the fewest sequences for target words among the experimental groups. For the control words however, they were similar to the other groups in all respects. When instances of two or three hits were examined in the post-intubation group, a high sequence rate was noted for colours, with 50% of patients producing a sequence of two and a further 12% a sequence of three, but this was not statistically significant.

EXPLICIT MEMORY

No anaesthetized patient had conscious memory of the tape being played or of its content. No patient had recall of events between the anaesthetic recovery and the recovery area. After category generation, the pre-induction group was asked, as the final question, for explicit recall of the auditory stimulus. The 84% hit rate for the group was identical to that obtained for category generation, although for individual patients there were inconsistencies.

RECALL IN RELATION TO ANAESTHETIC STATE

In summary, there was no evidence of implicit memory in the anaesthetized groups. The lowest enflurane concentration during the 5 min of tape playing was in the recovery group (mean 0.28%) and the highest was in the post-incision group (mean 1.05 (SD 0.2)%) (table 1). For the nitrous oxide and enflurane components, these means were estimated to be equivalent to 0.2 and 1.2 MAC, respectively. This calculation excludes the effect of the other anaesthetic constituents. No correlation was found, or relationship seen, between generation of the target words and anaesthetic state or stimulating events. The presence or absence of neuromuscular block had no effect on target word generation or on the mean end-tidal enflurane concentration; the mean enflurane concentration was 0.1% higher in paralysed compared with non-paralysed patients. The pre-induction group performed significantly better than the control group and all experimental groups in respect of hits, serial position and occurrence of sequences. It was also significantly better in the generation of target words compared with control words. Despite the low enflurane concentration, as intended, the recovery group showed a tendency to be worse in all respects than the other groups; this was not significant.

Discussion

Our results do not provide evidence that implicit memory can be established through priming of verbal material during general anaesthesia. However, we used a method which provoked a priming effect in conscious patients (pre-induction group) hence indicating that the task itself had the potential to detect priming if it had occurred. The results are consistent with negative studies\(^9\)\(^{10}\); failure to support previous positive studies\(^4\)\(^{7}\) may be explicable on one of several counts.

First, it could be that previous research made the methodological error of presenting stimuli throughout the procedure, hence increasing the chance of auditory stimuli occurring during unintentional light anaesthesia. Memory is expected to be absent for events from induction until surgery is completed, although approximately 0.5% of obstetric patients may have explicit memory of events where neuromuscular blocking agents are used during anaesthesia\(^6\); in non-obstetric anaesthesia this value may be nearer 1 in 3000.\(^7\) In paralysed patients, stimulating procedures during the surgical event are associated with lightening of anaesthesia.\(^8\)\(^9\) This study was designed to test the theory that implicit memory could be shown to occur at times when increased stimulation caused transitory arousal. Our method permitted examination of priming as a function of different stages of the surgical procedure, as did that of Eisele, Weinreich and Bartle.\(^1\) Our result is consistent with Eisele, Weinreich and Bartle\(^3\) but using a more advanced task and method. A later study, also using enflurane, showed a dose-related effect, with the early cortical components of auditory evoked responses (AER) being nearly abolished at an end-expiratory concentration of 1.22% enflurane in nitrous oxide.\(^2\) In our study, the end-expired enflurane concentration during tape-playing was less than 1.2% in 95% of the fully anaesthetized patients.

MAC is more difficult to calculate with this type of anaesthetic, as nitrous oxide and enfurane end-tidal concentrations can be measured but the contributions of premedication, induction agent and intra-operative analgesia are ignored. Nor are their effects equivalent in all groups; premedication and induction agent would have minimal effects in the recovery group but substantial effects in the post-induction group. The rationale behind our study was that lightening of the anaesthetic state might occur at times of stimulation despite apparently adequate clinical anaesthesia and MAC values. In the absence of a gold standard for anaesthetic depth, vapour concentration or MAC may permit some comparison between positive and negative studies in anaesthetized patients. In volunteer experiments, with a simpler and more rigid anaesthetic procedure, Dwyer and colleagues\(^2\) showed that the MAC-awake (the end-tidal concentration preventing voluntary response in 50% of subjects) for isoflurane was 0.38 times the MAC in healthy volunteers. Stress (i.e. surgical stimulation) may be expected to increase the concentration of isoflurane required to abolish implicit memory. In volunteers receiving 0.4% isoflurane, tetanic stimulation increased the coherent frequency and caused improved performance on memory tests, but these effects did not occur at 0.8% isoflurane.\(^2\) Matter-of-fact information presented at concentrations just above MAC-awake has shown no evidence of implicit memory. However, emotionally charged information required a higher concentration of anaesthetic, although still less than the MAC value, to suppress memory.\(^2\) Three of our four anaesthetized groups received the stimulus during stimulating events when arousal might occur.\(^1\)\(^8\)\(^9\) Despite this, implicit recall could not be detected in these subjects.

A second possible reason for failure to detect implicit memory may be that sample sizes are not...
large enough to detect an effect. It has been suggested that the reason for failure of a “perfectly conducted” anaesthetic could be related to a notional Gaussian distribution of sensitivity to anaesthesia which allows a period of wakefulness for a patient lying in the upper tail of the distribution.24 This would be relevant to the proposal that some individuals may be more sensitive to input, hence the sample size would need to be very large to include and detect them.

Third, it is conceivable that detection of implicit memory during anaesthesia may require even more sensitive tests, although the methodology of testing has been refined over the years; past methods included free-recall testing,13 cued-recall testing25 and word recognition tasks.26 A later modification avoided the recall dependence on the circumstances of learning.27 In our study using category generation, an attempt was made to optimize conditions for memory by avoiding benzodiazepine pre-medication,28 prefacing the test words with instruction to attend, and by presenting information at times of likely maximal stimulation. We deliberately avoided testing explicit memory first as this may impede the subsequent detection of implicit memory.7 Assessment of implicit memory was carried out within 36 h, as recommended by Merikle and Daneman.29

Future studies may consider several other aspects of methodology. Our study may be criticized for the lack of blinding as the investigators were aware of the words chosen, having been involved in the preliminary procedure to select them. However, this is unlikely to have affected our results. In selecting the target words for such studies, there should be no association between two target words (e.g. green and grape or orange and lemon), which might lead to their generation as a pair, regardless of whether they had both been primed. No particular association was evident in this study when sequences were examined. Methodologically, it might have been preferable to match priming targets for probability of their production in control conditions. However, their selection as relatively low frequency examples of the category was made to ensure that there was scope for priming effects to occur.

The relative ease with which categories are generated may need to be considered. In this population, colours were generated more easily than either fruits or vegetables. This was particularly interesting in the pre-induction group who were exposed to both colours and fruits and perhaps implicit central processing for storage and generation of this particular category. To reduce type 1 errors, categories which are generated less well should be used for target words rather than control words.

Future studies using category generation may also consider evidence that the modalities for presentation and retrieval of information should be the same, as implicit memory may be very sensitive to change in modality.30 State dependent learning (retrieval is dependent on the circumstances of learning being recreated at the time of recall) may have an effect on memory.31 Tone of voice32 during presentation of information to the patient (urgency/crisis) may provoke increased sympathetic stimulation but is still suppressed at less than 1 MAC for volatile agents.23

In this study, apart from successful priming in the pre-induction group, the only consistent trend (although not statistically significant) was the poorer recall of target words by the recovery group. This was surprising as, at the time of auditory stimulus, they appeared to be almost awake and were reacting to a tracheal tube or laryngeal mask in a vigorous manner. As this group was similar to the others in the production of control words, their inferior generation of target words may reflect the proposal that information presented at a time of noxious stimulus may be repressed.33

Finally, although there is a suggestion to the contrary,34 this study was based on the reasonable hypothesis that implicit memory is located in a continuum between explicit recall and no recall, as described by Jones and Konieczko35 and recently reiterated with substantial evidence.36 The final possibility remains that implicit memory does not occur in these circumstances. Despite the timing of the stimulus, implicit memory could not be located in this group of patients. Reliable replication of positive results in this field is proving difficult, but in the words of Altman and Bland37 “absence of evidence is not evidence of absence”. However, the occurrence of implicit memory in the adequately anaesthetized patient is proving difficult to substantiate.

References

Implicit memory during general anaesthesia