When does context influence recognition memory?

Duncan Godden and Alan Baddeley

Hewitt (1977) has distinguished intrinsic context, which is directly involved in the encoding of material and extrinsic context comprising such arbitrary features of the learning situation as environment of learning. While both types of context influence recall, with better performance when the original context is reinstated, recognition effects have been observed only with intrinsic context. The present study uses the contrast between the land and underwater environments to explore this apparent discrepancy. Subjects learned lists of 36 words either on land or under water, and subsequently tried to recognize them from a list of 72 words presented in either the same or the alternative environment. In contrast to an earlier recall study, no trace of context dependency was observed. Implications for the distinction between intrinsic and extrinsic context are discussed.

There has in recent years been a great deal of research into the role of contextual cues in memory. Perhaps the most clearly delineated position is that of Tulving (Tulving & Osler, 1968; Tulving & Thomson, 1973), who has proposed what he terms the encoding specificity principle. This assumes that a retrieval cue will be effective in prompting recall or recognition if, and only if, it was encoded with the relevant item during learning. Most studies on this issue have manipulated semantic context; for example Light & Carter-Sobell (1970) used words with more than one meaning (e.g. JAM) and showed that presenting such a polysemous word in one semantic context (JAM, strawberry), and testing in another (JAM, traffic) led to very poor performance. Other studies have shown that more subtle changes of semantic context may impair recall or recognition (Tulving & Osler, 1968; Barclay et al., 1974).

In a recent unpublished review of the literature on contextual effects in memory, Hewitt (1977) has drawn a distinction between intrinsic and extrinsic context. The term intrinsic context refers to aspects of a stimulus which are inevitably processed when the stimulus is perceived and comprehended. Examples would be the type of lettering in which a word is written, the voice in which an item is spoken, and typically the semantic characteristics of a word and its semantic context. Extrinsic context refers to those characteristics of the stimulus situation which are irrelevant to the processing of the stimulus itself; the colour of the walls of the room in which an experiment is carried out would be an example of an extrinsic contextual cue. While such cues have a less reliable influence on retention than intrinsic contextual ones, there is clear evidence that material learned in one environment is better recalled than in an alternative very different setting.

In a previous paper (Godden & Baddeley, 1975) we showed that the underwater environment allows a particularly clear demonstration of extrinsic context dependency, with divers who learned and recalled under water, or learned and recalled on dry land, remembering 46 per cent more than divers who learned in one environment and recalled in the other. Although the magnitude of the context dependency effect tends to be considerably smaller under the less dramatic manipulations of environment that are possible on dry land, it nevertheless presents a phenomenon of general interest from both a practical and a theoretical viewpoint.

There is reason to suspect that extrinsic and intrinsic contextual effects may differ in one important respect. Whereas intrinsic context has a powerful effect on both recall and recognition (e.g. Light & Carter-Sobell, 1970; Marcel & Steel, 1973; Watkins et al., 1976) all the positive evidence for an influence of extrinsic context comes from studies using recall, while the few recognition studies that have been performed appear to have produced uniformly negative results.
Baddeley et al. (1975) had divers learn passages of prose under water in either warm or cold conditions. On resurfacing, recall was required, followed by a task involving recognition of statements taken from the passages. Level of recall was considerably poorer than performance on a prior dry practice run, whereas recognition showed no comparable decrement. However, since the study was not set up to study context-dependency, the relevant controls were not included; nor was the dry practice passage strictly comparable with the underwater test passages. Davis et al. (1975) also obtained results suggesting that context dependency may be much stronger for recall than recognition. They tested divers three times on a task involving the recall and recognition of unrelated words. The first test was a practice run carried out on land, while the second and third were underwater tests in warm or cold water. In the recall test, subjects' performance dropped dramatically from a mean of 9.3 words correct on land to 4.8 in warm conditions and 3.0 in cold water. In the case of recognition, however, performance did not differ between dry land, with 23.8 words being recognized out of 30, and warm water where 22.7 words out of 30 were correctly recognized, although there was a small but significant effect of cold on performance. Once again however the experiment was not set up to study context dependency, and order of presentation is clearly a confounding factor.

A third source of consistent but inconclusive evidence comes from experiments on state-dependent memory, in which the subject's internal state is changed by means of a drug. A study by Goodwin et al. (1969) demonstrated clear state-dependency effects on a range of recall tasks when subjects were required to learn under the influence of alcohol and recall either drunk or sober. However, no context-dependent effect was found in a task involving the recognition of pictorial material. A study by Wickelgren (1975) again showed an effect of alcohol on learning, but no evidence of state dependency when performance was tested by recognition. Hence, although the available evidence is fragmentary and suggestive rather than conclusive, it does suggest that extrinsic context dependency effects may be avoided if memory is tested by recognition rather than by recall.

If one is predicting a negative result, it is clearly necessary to choose a situation in which powerful context dependency effects are known to occur. We therefore again opted to use the underwater environment which has been found to generate a much stronger effect than is typically obtained on land (Godden & Baddeley, 1975). Divers learned lists of words in both dry (D) and wet (W) conditions and subsequently were required to recognize them from a list containing an equal number of filler words, in either the same or the alternative environment. All divers performed under all possible conditions; DD (learn dry, recognize dry), DW, WW and WD.

**Method**

**Subjects**

Sixteen subjects, 12 male and four female members of a university diving club, were tested.

**Apparatus**

Five lists of words were constructed and subsequently recorded on tape (see Procedure). Each list consisted of 36 unrelated, two- or three-syllable words chosen at random from the Toronto word bank. The words were presented via a Diver Underwater Communication (DUC) set. This consists of a surface-to-diver telephone cable, terminating in a bone transducer, which, placed on the diver's mastoid, enables both surface-to-diver and diver-to-surface communication. The DUC set was modified such that taped material, monitored by the surface operator, could be presented directly to the subject using a cassette tape-recorder. A twin transducer on the set allowed two subjects to be tested during the same period. Weighted Formica boards, sealed with transparent Fablon, enabled subjects to record responses in pencil both on land and underwater. Subjects used standard SCUBA breathing apparatus and diving equipment of various designs dictated by personal preference.
Procedure

All instructions and stimuli for each experimental session were recorded on tape. Efficient auditory perception of stimuli by a submerged diver using SCUBA apparatus is seriously impaired by the noise of his breathing. The presentation of the material was therefore grouped, so as to allow the diver to adopt a comfortable breathing rate which did not interfere with his auditory perception. Thus, each list was presented in blocks of three words. Within each block, the words were spaced at 2 s intervals. Between each block, a 4 s interval enabled subjects to exhale, inhale and hold their breath in readiness for the presentation of the next block, and so on.

Each tape began with an explanation of this breathing procedure, followed by a ‘breathing pattern’ section, to ensure that subjects were breathing correctly and in rhythm before the first word of the list appeared. This section consisted of nine spoken presentations of the letter z, in three blocks of three, and with identical spacings to those of the words in the list itself. Immediately after each block of zs, subjects heard the command ‘breathe’. The presentation of the word list followed on naturally in rhythm, and the command to breathe was then dropped.

On each tape (one for each of the 16 condition/list combinations), the relevant list was presented twice. Between the first and second presentations, a gap of 10 s allowed subjects a short rest with unconstrained breathing. The second presentation was again preceded by the breathing pattern section.

To eliminate the complication of possible primary memory effects (Glanzer & Cunitz, 1966), the second presentation of each list was followed by 15 digits which subjects were required to copy at a rate of 2 seconds per digit. This was followed by the next instruction (e.g. ‘Ascend to the shore station’), and a 4 min delay. This delay occurred in all conditions and was necessary to enable subjects to comply safely with the relevant instruction. They were then required to recognize as many of the words as they could from a list containing the target words mixed randomly with an equal number of comparable filler words, also taken from the Toronto word bank. These were presented using typed lists attached to weighted Formica boards, which were water sealed. Responses were made by marking strips of PVC adhesive tape, attached to the boards alongside the lists, with a horizontal dash if a word was recognized, and a zero if it was not. They were instructed to work through the list serially, once only, making their decision about each word as it occurred. This took, on average, about 2 min.

The original 16 subjects were split at random into four groups of four. Prior to the first experimental session, all subjects underwent a practice session, comfortably seated around a table. During this they first practised the breathing technique, then the task itself, using a practice list.

Pairings of the remaining four lists L1, . . . , L4, with the four conditions, and the temporal orderings of the conditions for each of the four groups, were arranged according to a Graeco-Latin square design. Subjects experienced one condition per diving session, and the sessions were separated by approximately 24 hours. The design was such that each group experienced conditions and lists in different orders, that a given condition/list pair was never administered to more than one group, and that lists and conditions had equal representations on each experimental session.

Subjects in environment D (dry) sat by the edge of the water, masks tipped back, breathing tubes removed, and receivers in place. In environment W (wet), subjects dived to approximately 5 m, taking with them their Formica board and two pencils, and with their receivers in position. Subjects sat on the bottom with one arm looped round a heavy chain, and the session began after a verbal signal to the surface operator signified their readiness.

To control for the possibility of a ‘disruption’ effect, resulting from different amounts of activity between presentation and test in the different conditions (Strand, 1970), the following procedure was adopted. In the WW condition, subjects, after performing the digit copying task, surfaced to collect their response boards, then returned to their original submerged position to perform the recognition. In the DD condition, subjects were required to enter the water, dive, and return to the surface between digit copying and recognition. In addition, subjects learning ‘dry’ were required to get thoroughly wet and cold before the session began, to control for possible differential effects of cold. Testing took place at an open freshwater site near Cambridge, England.
Table 1. Recognition performance as a function of learning and test environment

<table>
<thead>
<tr>
<th>Learning environment</th>
<th>Test environment</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Dry</td>
<td>Wet</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hit rate</td>
<td>0.76</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>False positive rate</td>
<td>0.17</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>$d'$</td>
<td>1.87</td>
<td>1.89</td>
</tr>
<tr>
<td></td>
<td>Recall probability (previous study)</td>
<td>(0.36)</td>
<td>(0.24)</td>
</tr>
<tr>
<td>Wet</td>
<td>Hit rate</td>
<td>0.70</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td>False positive rate</td>
<td>0.19</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>$d'$</td>
<td>1.63</td>
<td>1.44</td>
</tr>
<tr>
<td></td>
<td>Recall probability</td>
<td>(0.23)</td>
<td>(0.32)</td>
</tr>
</tbody>
</table>

Figures in parentheses show recall probabilities from Godden & Baddeley (1975).

Results

Mean probability of correct recognition and positive probability are shown in Table 1. For comparison, recall scores from our earlier study are also presented. For context-dependency to be present, the effect of the environment on recognition should depend on the environment of original learning. This would be represented by an interaction between the cross-totals (DD + WW) and (DW + WD) in Table 1. Since the cross-totals were almost identical (1.44 and 1.45) there is clearly no trace of an interaction. Examination of the false positive rates suggests that the absence of an interaction is not due to the application of different criteria across the four conditions. This conclusion is reinforced if detections and false positives are combined to give a $d'$ measure; the slight difference in cross-totals (3.30; 3.52) is in the opposite direction to that predicted on the assumption that the recognition performance is context dependent. The only significant main effect was that of the environment of original learning, with more words recognized when learning took place on dry land than were recognized when it took place under water ($P < 0.01$). This could reflect either distraction caused by the less familiar underwater environment, or caused by possibly slightly noisier presentation conditions under water. The absence of an interaction is in marked contrast with the results of the previous study which used recall rather than recognition.

Discussion

Our results show no evidence for a context-dependent effect. As such they are consistent with the experiments described previously, and support the view that recognition memory is resistant to extrinsic context dependency. Why should this be so?

Let us consider first the two most prominent accounts of retrieval effects, namely the encoding specificity hypothesis and the list tagging hypothesis. An encoding specificity approach such as that advocated by Tulving & Thomson (1973) fits the recall data very well. If one assumes that environmental cues are encoded at the time of learning, then it follows that reinstating such cues should facilitate recall. It should, however, also facilitate recognition; indeed much of the strongest evidence supporting the encoding specificity hypothesis explicitly uses recognition rather than recall (Tulving & Thomson, 1973). Should one then conclude that the stimuli were not encoded during learning? If this is the
case, why then should reinstating the context enhance recall? It is far from clear how the encoding specificity can account for the observed results.

A list tagging hypothesis (e.g. Anderson & Bower, 1973) might attempt to account for the recall results by suggesting that contextual stimuli become associated with the list during learning, and serve as tags which help differentiate the material to be learnt from irrelevant material learnt elsewhere. reinstatement of the environmental context presumably primes that particular tag, making it more accessible and more usable. Once again, however, it is not clear why a similar effect should not be found with recognition.

One possibility is that the limit of performance in the recall task is set by retrieval, since there is good evidence in free recall to suggest that subjects typically learn much more than they can recall (see Baddeley, 1976, pp. 285-287 for a review). In the case of a recognition test, however, in which performance is much higher, it seems possible that the performance limit is set by the degree of initial learning. Such an explanation is possible but seems implausible since it assumes that a recognition test avoids all retrieval problems, and makes accessible all the information acquired during learning. Such a view is inconsistent with the demonstration by Brown (1964) that in a multiple-choice recognition test, if the first attempt is incorrect, then the second attempt has an above-chance probability of success. More recently, Tulving and his associates have demonstrated that subjects may be able to recall items they have failed to recognize (Tulving & Thompson, 1973), a result that suggests strongly that recognition does not eliminate the retrieval process, although it may influence it.

A further possibility, suggested by an anonymous reviewer, stems from the observation that in our experiment, extrinsic context is a feature which is common to all items in a list, whereas typically intrinsic context changes from item to item. This is certainly typically the case, and it would be of interest to separate these factors, although in practice this might prove very difficult to achieve. It is not clear, however, how the outcome of such a hypothetical study would help explain our present results. For both recall and recognition the context is constant throughout a particular list, so that any differences obtained seem unlikely to be attributable to failure to manipulate this particular variable.

A possible explanation is offered by Brown’s suggestion that “the primary difference between recall and recognition is that in recall access to the unit word code must be from the context code but in recognition access is guaranteed by the physical presence of the word itself” (Brown, 1976, p. 10). Since access from the context to the word is unreliable, anything which enriches this link by reinstating the original physical circumstances will enhance performance. In contrast to this, access to the word code from the printed word is much stronger and more reliable, and is unlikely to be facilitated by an extrinsic contextual cue.

Taken at face value, however, such an argument would suggest that recognition memory should be equally insensitive to the effects of intrinsic context, and that items which can be recalled should almost invariably be recognized. Both of these conclusions are inconsistent with available evidence (e.g. Light & Carter-Sobell, 1970; Tulving & Thomson, 1973). However, this apparent paradox ceases to be a problem if one makes the plausible assumption that intrinsic context directly influences what is learned. This is most obvious in the case of the polysemous words used by Light & Carter-Sobell (1970), where the semantic interpretation of the jam is completely different, when accompanied by the strawberry, from when it is presented with the cue traffic. The fact that both semantic meanings share a single graphemic and phonological code is irrelevant if learning occurs at the semantic level.

The same argument can be applied in the less extreme case such as that of Tulving & Osler (1968) who presented a word such as city with such separate cues as dirty and village. A concept like city is semantically very rich; the aspect of it that will be encoded, given the
cue dirty, is likely to include such associated features as garbage, traffic fumes and dust; in contrast, the city–village pairing suggests a quiet, clean, friendly enclave within a city.

A similar interpretation can be made of the experiments by Watkins et al. (1976), and by Winograd & Rivers-Bulkeley (1977), both of which show that the recognition of a photograph of a face may be influenced by the presence of a second face. At first sight, this might seem to be inconsistent with our claim that extrinsic context does not influence recognition. Typically, however, the subject is instructed to relate the faces, either explicitly as in the Watkins et al. instruction to judge the compatibility of the people depicted, or implicitly as in Winograd & Rivers-Bulkeley’s requirement that the subject make a decision as to the friendliness of the two people depicted. Since they were portrayed facing each other, such an instruction could easily be interpreted as requiring a judgement of the implied relationship between the two, hence again causing the encoding of one face to be influenced by the nature of the other.

In brief, we suggest that intrinsic context affects recognition memory because the context determines what is learned, and subsequently guides the subject back to the interpretation of the stimulus that occurred during acquisition. By definition, extrinsic context bears a purely arbitrary relationship with the material learned. As such, it does not determine the interpretation of the material, and hence can contribute nothing to the already powerful cues presented by the physical presence of the words to be remembered.

Acknowledgements

We are grateful to members of the Cambridge University Underwater Exploration Group who served as subjects, to Dr Karalyn Patterson for her comments on an earlier draft of this paper, and to the Medical Research Council for financial support.

References


Received 20 January 1978; revised version received 18 January 1979

Requests for reprints should be addressed to Duncan Godden, MRC Applied Psychology Unit, 15 Chaucer Road, Cambridge CB2 2EF.
Alan Baddeley is at the same address.