Shifting Attention in Cued Recall

Simon Dennis
University of Queensland

John K. Kruschke
Indiana University

In category learning, the order in which cases are presented affects how they are learned. Categories that are presented early are encoded in terms of their typical features, while categories that are presented late are coded in terms of their distinctive features. Kruschke (1996) suggested that learners shift their attention to the distinctive features in later learning to avoid interference from earlier cases. In this article, we show that the same principle applies in cued recall. Subjects consecutively studied two lists of word triples, using an anticipation procedure. The second list was composed of triples that contained one of the words from the first list. The pattern of cued recall results was the same as that observed in the category learning task, suggesting that a common mechanism of rapidly shifting selective attention underlies both situations. The results cannot be accounted for by global memory models, which have no mechanism for rapidly shifting selective attention. The paradigm provides a method for investigating selective attention in episodic memory that is not confounded with item characteristics.

Selective attention, that is, the tendency to focus on an item or set of items within a larger set, has long been thought to play a crucial role in episodic memory. However, establishing that selective attention is operative in a given paradigm, and finding ways to effectively manipulate selective attention, have proven difficult. Recent studies in the categorisation literature suggest a new methodology that could be used to provide converging evidence for the use of selective attention in episodic memory.

The paradigm emerged from work on how people utilise base rates, that is, relative frequencies of occurrence, in categorisation tasks (Gluck & Bower, 1988; Medin & Edelson, 1988). For example, Medin and Edelson used a simulated medical diagnosis situation, in which people learned to diagnose sets of symptoms as certain diseases. The subjects were given corrective feedback on each diagnosis and were trained to near-perfect performance. The diseases were separated into common and rare pairs, which appeared in random order with a 3:1 frequency ratio. Each case consisted of just two symptoms: one perfect predictor (denoted *PC* for the common disease and *PR* for the rare disease) and one imperfect predictor (denoted *I*), which was shared by the common and rare diseases. Medin and Edelson used three pairs of common and rare diseases, that is, six diseases and nine symptoms overall.

Subjects were subsequently tested with combinations of symptoms not encountered during training. When presented with the imperfect predictor (*I*) alone, they tended to choose the common disease. This preference is consistent with the base rates. When subjects were presented with the triplet of conflicting symptoms, $I + PC + PR$, they again tended to choose the common disease, although not as strongly. However, when presented with the pair of conflicting symptoms, $PC + PR$, subjects were more likely to choose the rare disease. This tendency goes against the base rates, and this pattern of results is called the *inverse base rate effect* (Medin & Edelson, 1988).

Kruschke (1996) argued that underpinning the inverse base rate effect (and apparent base rate neglect; Gluck & Bower, 1988) are two principles. First, all else being equal, subjects will learn and utilise base rate information. Second, during training, subjects insulate overlapping cases from each other by rapidly shifting attention to distinctive cues. A critical role of differential base rates is to make the common diseases occur earlier, on average, than the rare diseases. Consequently, subjects learn the common diseases before the rare diseases. Because the common diseases have no mutually overlapping symptoms, and because the overlapping symptoms in the rare diseases have not yet been learned, subjects learn about both symptoms, *I* and *PC*, of the common diseases. When subjects later learn the rare diseases, composed of symptoms *I* and *PR*, the learners shift attention away from the overlapping symptom *I*, toward the distinctive symptom *PR*, in order to avoid interference between the new, rare case, and the previously learned, common case. In effect, then, what subjects have learned is that symptoms *I* and *PC* share moderate associations with the common disease (*C*), but symptom *PR* is uniquely and strongly associated with the rare disease (*R*). When tested with symptom *I* alone, the associations and base rate work together, and subjects favour the common disease. When tested with the $I + PC + PR$ combination, the symptom associations are equivocal, but the base rates favour the common disease. When tested with the $PC + PR$ combination, the strong association from *PR* to *R* causes subjects chose the rare disease despite the base rates.

In Kruschke's (1996) argument, a critical role of the base rate is to determine when learning occurs. Consequently, the same result should obtain if some cue combinations are presented earlier than others. Kruschke (Experiment 2) provided support for this hypothesis. The design was similar to that described above. Instead of presenting items at different base rates, however, the study list was divided into two sets and different symptom–disease pairings were presented either early or late in training. The first list contained two $I + PE \rightarrow$

This research has been supported by Australian Research Council grant number A79701517 and by (USA) NIMH FIRST Award R29-MH51572. We would like to thank Michael Humphreys for his comments during the preparation of the manuscript.
Address for correspondence: Simon Dennis, School of Psychology, University of Queensland QLD 4072, Australia. E-mail: s.dennis@psy.uq.edu.au

E mappings, where I denotes an imperfectly predictive symptom and where PE denotes a perfectly predictive symptom of the early disease, E. The second list contained the same two mappings and two additional $I + PL \rightarrow L$ mappings, where I denotes an imperfect predictor which was that same as one of the predictors presented in the early mappings, PL denotes a perfect late predictor and L is a late target. The pattern of results from the test phase was the same as in the inverse base rate experiment. When the imperfect predictor, I, was presented alone, or when the triple ($I + PE + PL$) was presented, subjects favoured the early target, E. However, when the perfect early predictor (PE) and the perfect late predictor (PL) were presented together, subjects favoured the late target, L. The fact that this pattern of results was the same as the pattern in the inverse base rate experiment lends credence to the idea that the locus of the effect is in the learning history and that selective attention is critical to understanding the effect.

These principles of shifting selective attention were formalised by Kruschke (1996) in a connectionist model called ADIT, which accurately fitted data from four experiments. The model is described later in this article.

In transferring the category learning paradigm to the episodic memory domain, there are two important implications. The first implication is procedural: by manipulating the learning history of a cue, we can influence the amount of attention applied to the cue, independent of its idiosyncratic characteristics. This approach contrasts with paradigms that rely on item characteristics to manipulate attention. For example, the attention/likelihood theory (Glanzer & Adams, 1990) proposes that low frequency words are more distinctive and surprising than high frequency words, and therefore capture more attention. Consequently, one way of attempting to manipulate attention is to vary the frequency of the words employed. However, word frequency is correlated with many other item properties, such as number of associates, concreteness (Allen & Garton, 1968; McCormack & Swenson, 1972), degree of proactive inhibition (Gorman, 1961; McCormack & Swenson, 1972), structural or orthographic distinction (McCormack & Swenson, 1972; Zechmeister, 1972), and pre-experimental recency (Kinsbourne & George, 1974). Consequently, using frequency or other item characteristics to manipulate attention will generally permit a plethora of alternative explanations. Using the explicitly manipulated learning history of the item to redistribute attention circumvents this difficulty of confounded factors.

A second implication of the category learning results is theoretical. If comparable results obtain in episodic memory situations, the results oppose what would normally be predicted by strength/interference theories. To illustrate the difficulty that strength models encounter with the inverse base rate phenomena, consider the well known memory model, *search of associative memory* (SAM, Gillund & Shiffrin, 1984). SAM has been chosen because it shares the structure, typical of recent recognition models, that is responsible for the contrary prediction, and because it is explicit about the entire cued recall process.

The SAM model presumes that memory consists of a set of images. Memory retrieval involves applying a set of cues to activate one or more of these images. In cued recall, subjects are then assumed to engage in a sampling process in which they retrieve an image with a probability proportional to its degree of activation. Once an image has been retrieved, a recovery process extracts the target name from within the image. Our goal is to demonstrate that the model will incorrectly favour the early target E when the pair of cues, PE and PL, is presented in testing, after phased training on $I + PE \rightarrow E$ and $I + PL \rightarrow L$, as described above. To accomplish this goal, it suffices to show that both the sampling and recovery processes favour the early target.

In the sampling process, the activation of an image is calculated by multiplying together all the cue strengths, after they have been raised to a power that indicates the amount of attention applied to the cue. Formally, the activation of the j-th image is given by:

$$A(I_j \mid Q_1 \ldots Q_n) = \prod S(I_j, Q_i)^{W_i}$$

where A is the activation, I_j is image j, Q_i is cue i, $S(I_j, Q_i)$ is the strength from cue i to image j and W_i is the attentional weight of cue i. Attention is assumed to be a limited capacity system so the attentional weights are constrained to sum to one. The probability of sampling this image is determined by the magnitude of its activation relative to the sum of all activated images. This sampling probability, P_S, is specified formally by a form of the Luce (1959) choice rule:

$$P_S(I_j \mid Q_1 \ldots Q_n) = \frac{A(I_j \mid Q_1 \ldots Q_n)}{\sum_i A(I_j \mid Q_1 \ldots Q_n)}$$

In applying SAM to the inverse base rate paradigm, we assume that an image exists for each of the target items, and that the cues at test include the presented items and a context cue spanning both lists. (If the context cue spans only the second list, SAM has equal recall probability for E and L when presented with $PE + PL$, as can be derived from the discussion below.)

To demonstrate that the model will tend to incorrectly sample the early target when $PE + PL$ is presented we will show that the sampling probability of target E is greater than the sampling probability of target L, that is,

$$P_S(E \mid C,PE,PL) > P_S(L \mid C,PE,PL)$$

where E is the early target, L is the late target, C is a context cue spanning both study lists, PE is the perfect early cue and PL is the perfect late cue. Using the definition of the sampling probability, the inequality becomes

$$A(E \mid C,PE,PL) > A(L \mid C,PE,PL)$$

Now using the definition of activation, if we assume that the attention weights are equal, the inequality becomes

$$S(E \mid C)^{\frac{1}{3}}\ S(E \mid PE)^{\frac{1}{3}}\ S(E \mid PL)^{\frac{1}{3}} > S(L \mid C)^{\frac{1}{3}}\ S(L \mid PE)^{\frac{1}{3}}\ S(L \mid PL)^{\frac{1}{3}}$$

Taking the log of each side (which is monotonic increasing and hence preserves the order) we get:

$$\log S(E \mid C) + \log S(E \mid PE) + \log S(E \mid PL) > \log S(L \mid C) + \log S(L \mid PE) + \log S(L \mid PL)$$

Turning to the recovery probability we see that we get an expression that is similar in form. Gillund and Shiffrin (1984) define the probability of recovery as:

$$P_R(I_j \mid Q_1 \ldots Q_n) = 1 - \exp\left(-\sum_{i=1}^{M} W_i S(I_j \mid Q_i)\right)$$

We would like to prove that

$$P_R(E \mid C,PE,PL) > P_R(L \mid C,PE,PL)$$

which, by manipulations similar to those above, is equivalent to

$$S(E|C) + S(E|PE) + S(E|PL) > S(L|C) + S(L|PE) + S(L|PL)$$

Note that the inequalities for the sampling and recovery probabilities are similar in form except for the log transform. However, since the log transform is increasing monotonic, all of the arguments below apply to both the sampling and recovery probabilities.

Because the early target, *E*, was never studied with the perfect late predictor, *PL*, and because the late target, *L*, was never studied with the perfect early predictor, *PE*, it is reasonable that the strength of activation of *E* by *PL*, and the strength of activation of *L* by *PE*, are both very small and approximately equal, that is, $S(E|PL) \approx S(L|PE)$. Because the context spans both lists, it is reasonable that $S(E|C) \geq S(L|C)$ because the early targets are seen in this context twice as often as the late targets. Similarly, it is reasonable to suppose that $S(E|PE) \geq S(L|PL)$ because the perfect cues are always present when their respective targets appear, and the early target appears on more trials, overall. Consequently, by substituting these inequalities into the formulas derived above, it can be seen that both the sampling and recovery probabilities favour the early target *E* when *PE* + *PL* is presented, in violation of the observed results. Possible modifications of SAM are discussed near the end of this article.

EXPERIMENT: SHIFTING ATTENTION IN CUED RECALL

Our objective in this study was to establish that results analogous to the inverse base rate effect also occur in cued recall. Insofar as the attentional-shift explanation is correct, the implications of these results are twofold. First, this technique has promise as a means for manipulating attention in cued recall studies. Second, memory models should incorporate mechanisms for explaining such attentional shifts. We have chosen to replicate Experiment 2 from Kruschke (1996), but using a cued recall task rather than a categorisation task.

Method

Participants

Thirty-seven students volunteered for partial credit in an introductory psychology course at the University of Queensland.

Materials

Rather than employ sets of symptoms and diseases as in the categorisation experiments, we chose to use five-letter words with frequencies between 4 and 40 per million (Dennis, 1995) as both cues and targets. The appendix provides a complete list of the words we used. Our use of random words as both cues and targets, instead of symptoms and diseases, negates any concern that the implicit causality between diseases and symptoms might somehow influence the results (cf. Waldmann, Holyoak, & Fratianne, 1995). Random word stimuli are also quite typical of memory experiments.

Design and Procedure

The structure of the design was the same as that employed in Kruschke's (1996) Experiment 2. Each subject underwent two study sessions followed by a test. At each study presentation, the two cue words were presented beside each other at the top of the screen. Two seconds later the target word appeared at the bottom of the screen and remained for the next four seconds. Subjects were instructed to attempt to anticipate the target word covertly and to learn the cue target relationships for a subsequent test. This anticipation procedure has many precedents in the literature on memory research, for example, the classic works of Binder and Estes (1966) and of Underwood (1966).

For each subject, early and late training sets were constructed by randomly assigning words to the roles of imperfect predictor (*I*), perfect early predictor (*PE*), perfect late predictor (*PL*), early target (*E*), and late target (*L*). Four such sets were constructed. In the first study list, the imperfect predictor (*I*) and the perfect early predictor (*PE*) of each set were paired with the early target (*E*). In the second study list, the same four cue target combinations were presented as well as an additional four constructed from the imperfect predictor (*I*), the perfect late predictor (*PL*) and the late target (*L*) of each early/late set. Each combination of cues and target was repeated eight times, so that the early cue/target sets had been seen a total of 16 times by the end of the second list. Within each list, the order of presentations was randomised. For each presentation, the order of the cues was random; for example, an instance of *I* + *PE* could appear on the screen with cue *I* on the left and cue *PE* on the right, or with cue *I* on the right and cue *PE* on the left. Table 1 outlines the items presented in each of the study lists.

The test was self-paced. Cue sets, containing from one to three items, were presented, and the subjects were required to type in the word that seemed like the best target word. They were informed that on occasions there might be more than one correct target word and that when this was the case they were to choose the one that seemed the strongest. Both the order of the cues within each set and the order of presentation of the cue sets were randomised.

The test cue sets consisted of

1. Training cues: *I* + *PE* (e.g., Honey Razor), *I* + *PL* (e.g., Honey Plate)
2. Unambiguous cues: *PE* (e.g., Razor), *PL* (e.g., Plate)
3. Within-set ambiguous cues: *I* (e.g., Honey), *PE* + *PL* (e.g., Razor Plate) or all three cues, *I* + *PE* + *PL* (e.g., Honey Razor Plate)

Table 1
Design of Training Stimuli

		Example
Study List 1		
I1 + *PE1* → *E1*		Honey Razor Graph
I2 + *PE2* → *E2*	presented 8 times	Digit Album Shark
I3 + *PE3* → *E3*		Truck Fence Nurse
I4 + *PE4* → *E4*		Smoke Badge Chain
Study List 2		
I1 + *PE1* → *E1*		Honey Razor Graph
I2 + *PE2* → *E2*		Digit Album Shark
I3 + *PE3* → *E3*		Truck Fence Nurse
I4 + *PE4* → *E4*	presented 8 times	Smoke Badge Chain
I1 + *PL1* → *L1*		Honey Plate Reply
I2 + *PL2* → *L2*		Digit Trail Bonus
I3 + *PL3* → *L3*		Truck Teeth Shirt
I4 + *PL4* → *L4*		Smoke Jelly Clock

Note. To the left of the arrow are the cues. To the right of the arrow are the targets. *I* = imperfect predictor, *PE* = perfect early predictor, *PI* = perfect late predictor, *E* = early target, and *L* = late target.

4. Between-set ambiguous cues: *I + PEO* (e.g., Honey Album), *I + PLO* (e.g., Honey Trail), both perfect predictors, *PE + PLO* (e.g., Razor Trail), or *I + PE + PLO* (e.g., Honey Razor Trail). The *O* in the abbreviations for these cues indicates that the cue came from the "Other" set.

For the training, unambiguous and within-set ambiguous cue sets there were four observations drawn from each subject (one from each of the early/late sets) for a total of 148 observations. In the between-set ambiguous cases, only one observation was drawn from each pair of early/late sets to avoid the possibility of dependence. For instance, performance on the *PE*1 + *PL*2 cue set might be influenced by a previously occurring *PE*2 + *PL*1 cue set, because either *PE*1, *PL*2 or both may have already been retrieved at test. Consequently, only two observations could be drawn per subject and the total number of observations in each of these cases is 74.

Results

Table 2 shows the proportions of responses as a function of cues presented. The "no response" category indicates those occasions on which the subject clicked on the continue button or hit a carriage return without typing a response. The "other response" category indicates the occasions on which the word typed by the subject did not match any of the targets that had been paired with those cues in training.

Training Cues

We can assess the success of the training by considering the test responses to the cue sets that appeared at study (*I + PE* and *I + PL*). Subjects responded with the early target (*E*) when presented with *I + PE*, $\chi^2(1, N = 133) = 113.752$, $p < 0.001$,[1] and the late target (*L*) when presented with *I + PL*, $\chi^2(1, N = 135) = 91.267$, $p < 0.001$. Therefore, our use of eight repetitions of the anticipation procedure provided good levels of learning.

Unambiguous Cues

When presented with the early perfect predictor (*PE*), subjects chose the early target (*E*), $\chi^2(1, N = 113) = 97.567$, $p < 0.001$. When presented with the late perfect predictor (*PL*), subjects chose the late target (*L*), $\chi^2(1, N = 119) = 107.303$, $p < 0.001$. These results are unsurprising since these cues provided an unambiguous indication of the correct target.

Ambiguous Cues

When the imperfect predictor (*I*) is presented, however, either an early (*E*) or late (*L*) target would be a correct answer. In this case, subjects chose the early target (*E*) significantly more often than the late target (*L*), $\chi^2(1, N = 107) = 30.364$, $p < 0.001$. This finding is consistent with base rate information (over the entire study session), with the primacy of early training, and with the idea that during *I + PL* → *L* training, attention is shifted away from cue *I*.

Consistent with the Kruschke (1996, Experiment 2), when subjects were presented with both early and late perfect predictors (*PE + PL*), responses favoured the late target (*L*) over the early target (*L*), $\chi^2(1, N = 113) = 14.949$, $p < 0.001$.

Subjects were also presented with cue combinations from across sets of training words, for example, *PE*1 + *PL*2. Such combinations are collectively labelled *PE + PLO*, which denotes the perfect predictor of the early target from one set of words, combined with the perfect predictor of the late target from another set of words. Testing results showed that the *PE + PLO* cue combination also exhibited a trend towards the late (*LO*) targets, but the size of the effect was not as large and there were only half as many observations in this case, $\chi^2(1, N = 59) = 2.864$, $p = 0.091$.

Kruschke (1996, Experiment 2) found no significant preference for either the early and late target when subjects were presented with *I + PE + PL*. In contrast, we found evidence that subjects preferred the early target, $\chi^2(1, N = 125) = 6.722$, $p = 0.009$. Similarly, the *I + PE + PLO* cue combination produced a preference for the early target, $\chi^2(1, N = 62) = 28.452$, $p < 0.001$, as was the case in Kruschke's experiment. The strength of the preference was significantly greater for *I + PE + PLO* than for *I + PE + PL*, $\chi^2(1, N = 187) = 9.608$, $p = 0.002$.

Finally, in the *I + PEO* and the *I + PLO* case there are three valid responses. Either the subjects respond on the basis of the perfect predictor and produce *EO* or *LO* respectively, or they respond on the basis of *I*, in which case they could give an *E* response or an *L* response. In the *I + PEO* case, *EO* responses were significantly higher than *E* responses, $\chi^2(1, N = 52) = 6.231$, $p = 0.013$, and *E* responses were significantly higher than *L* responses, $\chi^2(1, N = 21) = 8.048$, $p = 0.005$. In the *I*

Table 2
Choice Proportion as a Function of Cue Combination

Cue Set Type	Cue Set	Choice proportion					
		E	*L*	*EO*	*LO*	No response	Other response
Training cues	*I + PE*	.865	.034	.014	.007	.014	.068
	I + PL	.081	.831	.007	.007	.007	.068
Unambiguous cues	*PE*	.736	.027	.014	.014	.020	.189
	PL	.020	.784	.007	.014	.014	.162
Within-set ambiguous cues	*I*	.554	.169	.007	.014	.014	.243
	PE + PL	.257	.541	.007	.007	.014	.176
	I + PE + PL	.520	.324	.020	.007	.000	.128
Between-set ambiguous cues	*I + PEO*	.230	.054	.473	.027	.014	.203
	I + PLO	.176	.081	.014	.554	.000	.176
	PE + PLO	.311	.000	.014	.486	.027	.162
	I + PE + PLO	.703	.014	.014	.135	.014	.122

$+ PLO$ case, LO responses were significantly higher than E responses, $\chi^2(1, N = 26) = 14.519$, $p < 0.001$, but there was no significant difference between the E and L responses, $\chi^2(1, N = 19) = 2.579$, $p = 0.108$.

Discussion

The results suggest that we have been successful in shifting attention in the cued recall domain, replicating Kruschke's (1996) categorisation results in an episodic memory task. In the $PE + PL$ case, subjects preferred the late target despite the fact that base rates favoured the early target. A straightforward explanation for the late target preference is that they paid more attention to the late perfect predictor when studying the $I + PL \rightarrow L$ stimuli, and we will examine the theoretical implications of this idea below. First, however, we make some methodological observations.

Methodological Observations

In several ways, the current experiment was more similar to typical memory experiments than categorisation experiments. First, the test was cued recall rather than a forced choice categorisation. Consequently, the field of potential responses was large in comparison to the four possibilities in Kruschke's (1996) paradigm. Second, the study procedure suggested covert anticipation rather than the overt anticipation required in the categorisation experiment. Third, the materials consisted of low to medium frequency five-letter words rather than symptoms and diseases, which demonstrates that the attentional effects do not depend on any implied causal relationships among the components. Finally, the number of cue–target relationships studied was double that in the original experiment. Despite these differences, the proportions of responses in each of the target categories are remarkably similar across the two experiments. Table 3 shows the results from this experiment, with the no response and other response categories removed. In brackets beside these figures are the results from Kruschke (Experiment 2). The correspondence is striking, especially in the key $PE + PL$ case. Figure 1 plots the data from our experiment against the data of Kruschke (Experiment 2).

Although our procedure clearly fits in the episodic memory domain, it did employ an anticipation method, which is more typical of categorisation experiments.[2] The anticipation method has been used in the memory literature previously (see Underwood, 1966 for a discussion), but differs from more current study tasks in that it establishes a clear cue–target relationship and involves multiple repetitions. Both of these factors may have been critical to establishing the effect. In particular, it may be important that the $I + PE \rightarrow E$ relationship be well learned to ensure that subjects notice the overlap of the imperfect predictor during late training. The role of the number of repetitions would be a fruitful avenue for future research.

A second methodological caveat that should be raised is the fact that during the cued recall test there was often more than one correct response, and so it was necessary to instruct subjects to choose the "stronger" of these responses. However, under these conditions, it is possible that decision factors other than the strength of the memory traces are playing a role. For instance, if subjects chose to discount targets that appeared in the first list, then they may favour the late target. Similarly, if subjects attempted to employ the proportion of shared context between cues and targets in the decision procedure then they might favour the late target, because the $PL \rightarrow L$ mapping occurred only in the second list.[3] In both cases, however, it would also be necessary to explain why this advantage for later items does not also occur when cue I is presented alone.

Theoretical Implications

In this section, we examine the theoretical impact of our results. We start by demonstrating how Kruschke's (1996)

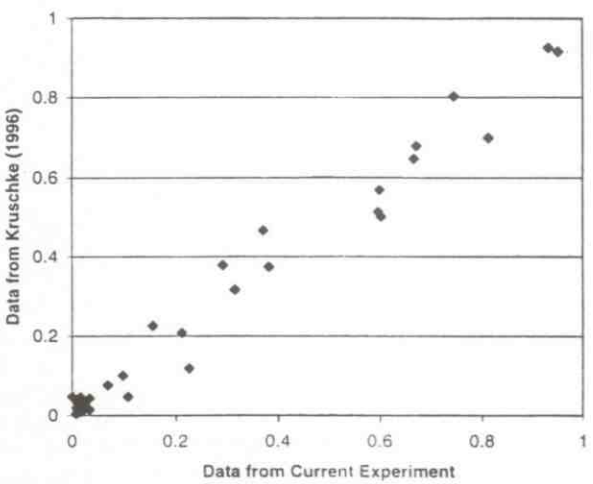

Figure 1
Correspondence of our cued-recall data with Kruschke's (1996, Experiment 2) categorisation data.

Table 3
Comparison of Choice Proportions in the Current Experiment and Kruschke (1996) experiment 2

Cue set type	Cue set	*E*	*L*	*EO*	*LO*
Unambiguous	*PE*	.932 (.925)	.034 (.014)	.017 (.038)	.017 (.024)
	PL	.025 (.028)	.951 (.915)	.008 (.019)	.016 (.038)
Within-set ambiguous	*I*	.745 (.802)	.227 (.118)	.009 (.033)	.018 (.047)
	PE + PL	.317 (.316)	.667 (.646)	.008 (.019)	.008 (.019)
	I + PE + PL	.597 (.514)	.372 (.467)	.023 (.014)	.008 (.004)
Between-set ambiguous	*I + PEO*	.293 (.379)	.069 (.076)	.603 (.502)	.034 (.043)
	I + PLO	.213 (.208)	.098 (.100)	.016 (.014)	.672 (.678)
	PE + PLO	.383 (.374)	.000 (.047)	.017 (.009)	.600 (.569)
	I + PE + PLO	.813 (.698)	.016 (.047)	.016 (.028)	.156 (.226)

Note. Proportions in brackets are those from "Base Rates in Category Learning", by J.K. Kruschke, 1996, *Journal of Experimental Psychology: Learning, Memory and Cognition*, 22(1), pp. 3–26.

ADIT model of categorisation can account for our cued recall results. Then we consider what modifications could be made to the global matching models (and in particular the SAM model used earlier) to encapsulate shifting attention. Finally, we comment on the value of maintaining separate categorisation and memory modelling literatures.

Attention to distinctive input (ADIT). ADIT is perhaps the simplest possible connectionist implementation of the idea that predictive error drives both the shifting of attention and the changing of associative weight. When a set of cues is presented with a target, the model computes the discrepancy between its anticipated target and the actual target, and then it performs two actions to reduce this error. First, it shifts attention away from cues that cause error, and it shifts attention toward cues that reduce error. Second, it changes association weights, between cues and targets, to reduce any remaining error.

When applied to the training sequence of our experiment, ADIT behaves as follows. When learning the first list, in which there are cases of $I + PE \rightarrow E$, ADIT attends to both cues I and PE because they do not conflict with any other instances in the first list. Thus, ADIT learns moderate associative weights between cue I and target E, and between cue PE and target E. When learning instances of $I + PL \rightarrow L$ during the second list, ADIT shifts attention away from cue I because it causes error. Cue I causes error because it is already associated with target E. The shift of attention, away from cue I, protects and preserves the previously learned association from I to E. The shift also facilitates more rapid learning of target L because the shift prevents interference from the ambiguous cue I. The associative weight from PL to L grows to a relatively large value because cue PL must, by itself, anticipate target L, whereas cues I and E share anticipatory duties for target E. After learning, when tested with cue combination $PE + PL$, the strong association from PL to L overpowers the moderate association from PE to E, and the target L is preferred. When tested with cue I alone, target E is anticipated by the moderate association from I to E, which was learned during the first list and preserved by attention shifts during the second list. When tested with cue combination $I + PE + PL$, either target could be weakly preferred, depending on the specific quantitative trade-offs in the model.

Kruschke (1996) showed that ADIT provided accurate quantitative fits to categorisation data. Given the dramatic similarity between the results of our current experiment and the results of Kruschke's (1996) Experiment 2 (recall Table 3 and Figure 1), it is plausible that ADIT would also provide a good fit to our new data. To address the cued-recall paradigm, as opposed to the forced-choice categorisation paradigm, ADIT needs an additional category for "other" responses. The same formalisms used for the original model, as detailed by Kruschke (1996), apply to this situation. Only one new parameter is needed, namely, the constant activation of the "other" target.

A brief review of the formalisms in ADIT is now presented, followed by the best fitting choice proportions. ADIT consists of a connectionist network with a set of input nodes that represent cues, and a set of output nodes that represent targets. The activation of a cue node is 1 when the cue is present, and the activation is 0 otherwise. Initially on each trial, all present cues are equally attended to, with attention strengths, α_i, normalised such that

$$\sum_i \alpha_i^P = 1,$$

where P is a freely estimated attentional capacity constant. Activation propagates from cue nodes to target nodes via weighted connections, such that the activation of target node k, is given by

$$a_k^{tar} = \sum_i w_{ki} \alpha_i a_i^{cue}$$

where w_{ki} is the associative weight from the i-th cue node to the k-th target node. We assume that the activation of the target node that represents "other" responses is a fixed and freely estimated constant, a_{oth}^{tar}. The probability of recalling target k is given by the Luce (1959) choice rule,

$$p_K = \frac{\exp(\varphi a_k^{tar})}{\sum_k \exp(\varphi a_k^{tar})}$$

where φ is a freely estimated scaling constant, and k ranges over all of the target nodes, including the "other" node. (Unlike the original ADIT model, base rate bias is not included in the present application, because the base rates of the targets, in the second list, were equal.)

When the correct target is provided during training, the target nodes are provided with "teacher" values, t_λ, which are 1 for the target that is present and zero for the other targets. The first consequence of the target being provided is a rapid shift of attention to reduce the error,

$$E = .5 \sum_\lambda (t_\lambda - a_k^{tar})^2,$$

where the sum is taken over all nodes but the "other" node. The attention is shifted in the direction (opposite) of the gradient of the error with respect to attention, which yields the following formula for attention shifts:

$$\Delta \alpha_i = \lambda_\alpha \sum_k (t_k - a_k^{tar}) w_{ki} a_i^{cue}$$

where λ_α is a freely estimated constant of proportionality called the attention shift rate. After attention is shifted, it is renormalised, and activation is repropagated to the target nodes, where the remaining error is computed. The association weights are then adjusted by gradient descent on the remaining error, which yields the following formula for associative weight changes:

$$\Delta w_{ki} = \lambda_w (t_k - a_k^{tar}) \alpha_i a_i^{cue}$$

where λ_w is a freely estimated constant of proportionality called the *association weight learning rate*.

The model was fitted to the data of Table 2, with the "no response" values deleted. The predictions of the model were determined from the mean choice probabilities of 200 randomly generated simulated subjects. Best fitting parameter values were estimated by minimising the log-likelihood statistic

$$G^2 = 2 \sum_c f_c \ln\left(\frac{f_c}{\hat{f}_c}\right),$$

where the sum is taken over all cells of the data table, where f_c is the observed frequency of responses in cell c, and where $\hat{f}_c$ is the predicted frequency of responses in cell c. To adjust for possible non-independence of responses from repeated tests given to a single subject, the frequencies were determined by first computing the proportion of each choice for each test case, and then multiplying by the number of subjects. As there were five possible responses (*E*, *L*, *EO*, *LO*, and *Other*) and 11

test cases, there were 55 frequencies fitted by the model. The choice frequencies for each case must sum to the number of subjects, however, so there were 44 degrees of freedom in the data. The model has five free parameters, and therefore the value of G^2 is compared with a chi-square distribution with 39 degrees of freedom.

Table 4 shows the best fitting choice proportions of ADIT, obtained from parameter values of $\varphi = 5.04$, $a^{tar}_{oth} = 0.222$, $\lambda_\alpha = 1.71$, $\lambda_w = 0.195$, and $P = 3.03$, which yielded $G^2_{(39)} = 17.37$, a value that suggests that the model cannot be rejected on statistic grounds alone. (For these parameter values, the root mean squared deviation was 0.0289, but this was not minimised.) ADIT shows all the effects seen in the human preferences; in particular, for cue *I*, ADIT recalls target *E* much more than target *L*, and for cue set *PE* + *PL*, ADIT recalls target *L* substantially more than target *E*.

The importance of attention shifts is highlighted by observing the model's best fit when it is restricted to learning without attention shifts. When the attention shift rate is fixed at zero, the best fit rises from $G^2_{(39)} = 17.37$ to $G^2_{(40)} = 53.25$, a highly significant increase. More importantly, the model without attention shifts shows no preference for target *L* when tested with cues *PE* + *PL* (model predictions of .300 and .277 for *E* and *L* respectively, compared with .260 and .548 by humans), and the model without attention shifts shows a preference for target *E* when presented with *PE* + *PLO*, contrary to human behaviour (model predictions of .493 and .328 for *E* and *LO*, compared with .319 and .500 for humans). Thus, in ADIT, attention shifts are critical for reproducing the preferences shown by human learners.

The success of the model, with its reliance on attentional shifts, provides additional evidence that attentional shifts may indeed be occurring in human learners during encoding of the lists. ADIT also provides a mechanism by which the attentional shift is generated, and this mechanism has the same motivation as the mechanism that generates associative weight changes. The motivation is simply rapid reduction of error when trying to anticipate the target. Attentional shifts in ADIT are essentially a mechanism for reducing interference between previously learned associations and newly demanded responses. When learning I + PL → L, after already having learned that I + PE → E, shifting attention away from cue I protects previously learned associations from backward interference with the newly demanded target, and shifting attention away from cue I also eases learning of the new association by reducing forward interference with previously learned associations. When rapid learning is important, shifting attention during encoding is beneficial.

Modifying the global matching models. In the introduction, we worked through a simple application of the SAM model to this domain, demonstrating that it predicts an inappropriate preference for target *E* when the *PE* + *PL* cue set is presented. There are, however, alternative assumptions that would allow SAM (and, with analogous modifications, other global matching models) to account for the results. First, it could be assumed that the retrieval context cue, *C*2, should span only the second list, and that $S(L \mid C2) > S(E \mid C2)$ because more attention is paid to the late targets at study, as they are novel. In addition, to correctly predict the *I* + *PE* + *PL* results, the imperfect predictor must continue to favour the early target. Consequently, attention must be focused on the *PL* cue when the *I* + *PL* cue combination is presented. This explanation invokes notions of attention in a similar fashion to ADIT, but includes no specific mechanism for how attention shifts.

A second modification of SAM involves the attention weights at test. If subjects were more inclined to attend to the perfect late predictors at test, perhaps because they are more salient in the second list, then $S(L \mid PL)^{Wt}$ could be larger than $S(E \mid PE)^{We}$, and the late target could be favoured when tested with *PE* + *PL*. When the *I* + *PE* + *PL* cue set was presented, however, attention might also be distributed to the additional *I* cue which favours the early target, and so the effect could reverse. Both this explanation and the ADIT explanation rely on a shift of attention to the perfect late predictor. In ADIT, the shift occurs at study, whereas in this modification of SAM it happens at test. Again, however, this modification of SAM includes no mechanism that governs the shift of attention.

At a broader level, the similarities, outlined above, between the categorisation and cued recall results bring into question the utility of viewing categorisation and episodic memory as separate domains of interest. Categorisation testing is functionally equivalent to a forced choice recognition test. Although the paradigms differ somewhat on the nature of variables

Table 4
Choice Proportions of ADIT (A.) and Humans (H.)

Cue set type	Cue set	Choice									
		E		*L*		*EO*		*LO*		Other	
		A.	H.	A.	H.	A.	H.	A.	H.	A.	H.
Training cues	*I* + *PE*	.923	.877	.012	.034	.007	.014	.007	.007	.051	.068
	I + *PL*	.056	.082	.810	.837	.015	.007	.015	.007	.105	.068
Unambiguous cues	*PE*	.742	.752	.014	.028	.027	.014	.027	.014	.190	.193
	PL	.006	.021	.829	.795	.018	.007	.018	.014	.128	.164
Within-set ambiguous cues	*I*	.563	.562	.118	.171	.035	.007	.035	.014	.249	.247
	PE + *PL*	.210	.260	.461	.548	.036	.007	.036	.007	.256	.178
	I + *PE* + *PL*	.512	.520	.343	.324	.016	.020	.016	.007	.113	.128
Between-set ambiguous cues	*I* + *PEO*	.271	.233	.078	.055	.420	.479	.018	.027	.212	.205
	I + *PLO*	.226	.176	.065	.081	.010	.014	.522	.554	.176	.176
	PE + *PLO*	.326	.319	.014	.000	.010	.014	.486	.500	.164	.167
	I + *PE* + *PLO*	.748	.712	.016	.014	.005	.014	.154	.137	.076	.123

Note. Human data are the same as Table 2, but with "no response" removed from the total.

manipulated and typical design parameters (such as number of learning trials), the maintenance of separate modelling literatures seems unwarranted.

CONCLUSIONS

The role of selective attention in episodic memory has been controversial (Clark & Gronlund, 1996). Previous attempts to establish the important of selective attention have either been confounded with item characteristics, or confounded with strength/interference. The results described in this paper establish the use of selective attention during encoding in cued recall free from these confounds. Furthermore, the close correspondence between our results and those outlined by Kruschke (1996) suggest that the same attentional mechanism underlies both episodic memory and categorisation. Models of memory need to include mechanisms for shifting attention.

Footnote

1. These chi-square values were computed from the raw frequencies of the responses with a null hypothesis that the number of observations in both categories is equal. Individual participants contributed more than one response to Table 2, which raises the possibility that independence was violated, thereby invalidating the probability values corresponding with the chi-square statistics. However, most of the chi-square values in our tests are very large, so that even highly conservative adjustments of the chi-square value (Wickens, 1989, p. 28) lead to the same conclusions.
2. Thanks to Stephan Lewandowsky for raising this point.
3. Thanks to Scott Gronlund for raising this point.

APPENDIX

Words Used in the Experiment

Album	Badge	Bonus	Chain	Clock
Digit	Fence	Graph	Honey	Jelly
Nurse	Plate	Razor	Reply	Shark
Shirt	Smoke	Teeth	Trail	Truck

REFERENCES

Allen, L.R. & Garton, R.F. (1968). The influence of word knowledge on the word frequency effect in recognition memory. *Psychonomic Science, 10*, 401–402.

Binder, A., & Estes, W.K. (1966). Transfer of response in visual recognition situations as a function of frequency variables. *Psychological Monographs: General and Applied, 80*(23), 1–26.

Clark, S.E., & Gronlund, S.D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin and Review, 3*(1), 37–60.

Dennis, S. (1995). The *Sydney Morning Herald* word database. *Noetica: Open Forum* [On-line], *1*(4). Available: http://psy.uq.edu.au/CogPsych/Noetica/

Gillund, G., & Shiffrin, R.M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91*(1), 1–67.

Glanzer, M., & Adams, J.K. (1990). The mirror effect in recognition memory: Data and theory. *Journal of Experimental Psychology: Learning, Memory and Cognition, 16*(1), 5–16.

Gluck, M.A., & Bower, G.H. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General, 117*, 227–247.

Gorman, A.N. (1961). Recognition memory for names as a function of abstractness and frequency. *Journal of Experimental Psychology, 61*, 23–29.

Kinsbourne, M., & George, J. (1974). The mechanism of the word frequency effect on recognition memory. *Journal of Verbal Learning and Verbal Behavior, 13*, 63–69.

Kruschke, J.K. (1996). Base rates in category learning. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 22(1), 3–26.

Luce, R.D. (1959). *Individual choice behavior: A theoretical analysis*. New York, NY: Wiley.

McCormack, P.D., & Swenson, A.L. (1972). Recognition memory for common and rare words. *Journal of Experimental Psychology, 95*, 72–77.

Medin, D.L., & Edelson, S.M. (1988). Problem structure and the use of base-rate information from experience. *Journal of Experimental Psychology: General, 117*, 68–85.

Underwood, B.J. (1966). *Experimental psychology*. New York, NY: Meredith .

Waldmann, M.R., Holyoak, K.J., & Fratianne, A. (1995). Causal models and the acquisition of category structure. *Journal of Experimental Psychology: General, 124*(2), 181–206.

Wickens, T.D. (1989). *Multiway contingency tables analysis for the social sciences*. Hillsdale, NJ: Erlbaum.

Zechmeister, E.B. (1972). Orthographic distinctiveness as a variable in word recognition. *American Journal of Psychology, 85*, 425–430.