

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* →

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:

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

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:

$$P_S(I_j | Q_1 \dots Q_n) = \frac{A(I_j | Q_1 \dots Q_n)}{\sum_i A(I_i | Q_1 \dots Q_n)}$$

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,

$$P_S(E | C, PE, PL) > P_S(L | C, PE, PL)$$

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

$$A(E | C, PE, PL) > A(L | C, PE, PL)$$

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

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

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

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

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:

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

We would like to prove that

$$P_R(E | C, PE, PL) > P_R(L | C, PE, PL)$$

