

## The Functionalist Agenda in Memory Research

*James S. Nairne*

One of the benefits of working in applied settings is that problems tend to be well defined. It is relatively easy to measure progress, and analytical techniques can be objectively evaluated. If the charge is to help a banker remember the vault code, it is not difficult to tell whether the intervention is working—the vault either opens or remains closed. For the laboratory researcher, however, problems are often ill defined. There can be no clear starting point, well-stated goal, or simple way of evaluating or marking progress.

Consider the study of pure memory, an enterprise widely pursued by researchers in cognitive psychology. Most memory researchers focus on a particular memory system or process, such as episodic, semantic, or implicit memory, and they conduct experiments to (a) isolate its underlying mechanisms and (b) determine its parameters of operation. The goal is to analyze the system's structure and its component parts in much the same way that a chemist might analyze a chemical compound by breaking it down into simpler elements. Attention is rarely given to the system's function at this point, largely because understanding function is presumed to depend on understanding structure. After all, how can we determine the function of a system unless we first understand the system itself? It is necessary to isolate the critical components, along with some rules for their interaction, and then—perhaps—the adaptive role that the system plays in cognition can be specified.

In this chapter, I discuss some of the implications of this widely practiced structuralist, or nonfunctional, approach to the analysis of memory. To begin, as noted above, problems crafted within a structuralist framework tend to be ill defined. When we set out to study *implicit memory*, it is difficult to gauge progress, or measure success, because the objective is unclear. The components of an implicit memory system are unknown; consequently, there is no way of determining when, or if, the system has been fully described. Researchers often end up studying tasks as a result, such as paired-associate learning or word-fragment completion, because task performance is easy to evaluate objectively. The trouble with this focus, however, is that the link between the studied tasks and the true memory system of interest can be tenuous or, more likely, inadequately specified.

Second, there are compelling reasons to believe that our memory systems are functionally designed. Our capacity to access and use the past did not develop in a vacuum; instead, our memory systems evolved to help us solve particular problems, adaptive problems that arose in our ancestral past. Without a functional perspective, as a result, it will be extremely difficult to determine how any mnemonic system works (or we can easily be led astray). As a starting point, we should assume that the structural components of the system exist, and work the way they do, because they are design features—that is, they contribute to solving a problem that is regularly faced by the organism (Klein, Cosmides, Tooby, & Chance, 2002; Tooby & Cosmides, 1992). Note that structure from this perspective is a by-product of function rather than the other way around: Nature designs, or selects, particular structural features because they aid in solving a problem faced by the species (Dennett, 1995).<sup>1</sup>

Finally, as I discuss in the main body of the chapter, ignoring the primacy of function in our analyses can lead to principles, or theoretical positions, that are misleading or wrong. Two such widely held principles are considered here: (a) Memory, especially short-term memory, depends directly on the activation of a trace, which, in turn, decays over time; and (b) memory, especially long-term memory, depends directly on the similarity or match between retrieval cues and traces, as encoded. Note that neither of these principles is placed in any kind of functional context; there is no consideration given to a central problem that the memory system might be attempting to solve, other than a vague reference to “memory.” As a consequence, we are left with principles that generate equivocal predictions and, in fact, violate general dictums of memory theory in some cases (see Nairne, 2001, 2002a, 2002b). Before developing these points in detail, I begin by addressing the question that sits at the heart of any functional analysis: What is memory for?

### What Is Memory For?

Everyone agrees that the ability to store, recover, and use the past serves an adaptive role. But did our memory systems evolve primarily to solve the problem of reproducing the past or to solve other problems? Some have argued that memory’s primary function is to help us use the past, in combination with the present, to decide on an appropriate plan of action (e.g., N. H. Anderson, 1996; Glenberg, 1997). In some respects, designing a system literally to reproduce the past makes little adaptive sense. A past event can never occur in exactly the same way again, nor can it be perceived in a like manner. In the words of

---

<sup>1</sup>Not all structural components are necessarily adaptations, that is, components that have been sculpted by the mechanisms of natural selection. It is still reasonable to assume, however, that their role in remembering is functionally designed, regardless of whether they are adaptations, exaptations (i.e., co-opted components that evolved for different reasons), or even spandrels (i.e., components that by themselves are not adaptations but are linked in some way to other evolved traits; e.g., Gould & Vrba, 1982). Moreover, my use of the term *designed* is not meant to imply the existence of a preexisting plan but rather refers to the process through which components conform to prevailing environmental demands.

William James, identical experiences require “an unmodified brain. The organ, after intervening states, cannot react as it did before” (James, 1884, p. 11).

In addition, granted that our memory systems are functionally designed, the stored components might well have little meaning outside of a particular goal context. For example, stored experiences could simply feed into an inferential “whole” that renders the experiences themselves unimportant, or changes them in such a way that they become unrecognizable as specific experiences. Trait judgments are easily remembered and applied—for example, whether we think an individual is likable—but they seem not to depend on recovery of the specific episodes on which they are based (N. H. Anderson, 1996; Klein et al., 2002). Scripts and schemas regularly guide remembering, such as when we attempt to recall our latest visit to a fast-food restaurant, but our memories often bear little resemblance to any particular episode or experience (Bower, Black, & Turner, 1979).

One can even imagine conceptualizations of memory in which the individual components are never actually stored. Experience might merely change affordances, or the range of possible interactions that we can have with the environment (Glenberg, 1997). Cutting off the leg of a chair changes how we can interact with it in the future, regardless of how or when the original removal took place. Similarly, experience might alter our perceptions, or the likelihood of certain responses, without reference to the situations that produced the change. The original events need not be reproduced, or remembered accurately, to exert lasting effects on our cognitions or behaviors. Present behavior can reflect the past, a true function of memory, without a lasting record of the event that precipitated the change.

At the same time, reproduction is not completely useless. We need to remember to take medication at a particular time, call to mind the name of a colleague, and remember necessary telephone numbers or passwords. Our ancestors in the Pleistocene era certainly needed to remember the specific locations of food, water, and competing groups. Nothing inferential would be appropriate in these situations; it is the literal location or sequence of numbers that is needed to accomplish the goal. Reproduction thus becomes an adaptive outcome, and we can assume that our memory systems contain functionally designed mechanisms appropriate to this end. (This is a comforting conclusion given that the vast majority of laboratory research over the past century has been directed at studying literal reproduction [Koriat, Goldsmith, & Pansky, 2000].)

It is important, however, to note that any functional design for reproduction needs to contain mechanisms for ensuring that the act of reproduction is context dependent. Passwords and telephone numbers are appropriately remembered only within specific environments, and there are costs associated with inappropriate retrievals (J. R. Anderson & Milson, 1989). Reproduction thus becomes a matter of discriminating where and when to produce a response based on information available in the present. At any given moment, a wide range of responses is available (e.g., each of us must now remember multiple user IDs and passwords)—the task is to select the response that is most appropriate for that context. Recognizing the context-dependent nature of reproduction is crucial because it implies that our theories of remembering cannot simply focus

on the characteristics of a response, or a memory trace; instead, the focus should be on conditional response selection: For a given environment, how is the most appropriate response, or memory, selected? The outcome of the selection process will depend on the features present in the situation (the retrieval cues) as well as on the characteristics of any stored information (e.g., Toth & Hunt, 1999; Tulving, 1983).

### **Activation, Decay, and Remembering**

With this functional analysis of reproduction in mind, I turn now to the oft-argued proposal that memory, particularly short-term memory, depends on the activation of a memory trace, which, in turn, decays over time. Note that reproduction is tied directly to a property of the trace in this case—activation—and no reference is made to context, or to any kind of response selection process. Activation alone determines retention and, although there may be situations that increase or decrease activation levels, reproduction itself is assumed to be context independent.

It is interesting to note that short-term memory may be one instance in which reproduction is not context dependent; that is, information sits at the focus of awareness and, as such, is instantly available. In William James's original conception of primary memory, items did not need to be retrieved because they were, in effect, already in a state of retrieval: "an object of primary memory is not thus brought back; it was never lost; its date was never cut off in consciousness from that of the immediately present moment" (James, 1890/1983, p. 608). The Jamesian view maps easily onto the concepts of activation and decay. One can conceive of information receding slowly into the past, hastened along by a decay process, until it drops below a threshold of awareness and instant availability.

Modern conceptions of working memory adopt the Jamesian idea but package it in a more modularized architecture. For example, Cowan (2001) distinguished between activated information in memory and a smaller, more restricted, subset of information that sits at the focus of awareness. The latter suffers from capacity limitations (roughly four chunks), and its contents can be immediately accessed—that is, the items reside in a "retrieval-free" zone. As Kintsch and others have noted, it is common for working memory theorists to assume "that information 'in' working memory is directly and effortlessly retrievable" and that retrieval from working memory, unlike long-term memory, is "not cue dependent" (Kintsch, Healy, Hegarty, Pennington, & Salthouse, 1999, p. 414). Consider Cowan's (2001) commentary on the finding that proactive interference (PI) sometimes fails to occur in immediate recall contexts: "This presumably occurs because four or fewer items are, in a sense, already retrieved; they reside in a limited-capacity store, eliminating the retrieval step in which PI occurs" (p. 103).

Baddeley's popular working memory model (Baddeley, 2000; Baddeley & Hitch, 1974) has a similar flavor, although in Baddeley's conception capacity limitations arise from trade-offs between decay and processes devoted to maintaining above-threshold activation levels (e.g., rehearsal). Active traces are

maintained in loops and stores and are forgotten whenever reactivation fails to counteract the relentless ravages of decay. Juggling is an apt metaphor: Items can be maintained aloft to the extent that each can be caught and retossed before gravity drives them into the ground. Here, capacity limitations are not a property of air (e.g., a store or a resource) but rather are determined by the balance between retossing (activation) and the downward pull of gravity (decay). Unlike Cowan, Baddeley avoids strong claims about the immediate availability of stored items—that is, items may need to be retrieved—but retention in his model remains closely tied to activation and forgetting is closely tied to decay.

From a functional perspective, it seems adaptive to possess a system that keeps subsets of information immediately available, perhaps as a necessary ingredient for language processing or online problem solving. In the working memory system, an articulatory rehearsal process might have developed to help us prolong or maintain information in a form suitable for producing and comprehending spoken language. However, to link an item's mnemonic status solely to its activation level, or to the simple passage of time, makes virtually no adaptive sense. Records of the recent past have value in some situations and not others, regardless of their activation status. When asked to remember a just-presented telephone number, it is important to maintain a literal record of the immediate past. However, one can easily imagine situations, perhaps even the majority of situations, in which such a maintenance process would lead to considerable clutter. Functional memory is a matter of target selection, picking and choosing the most appropriate response given the current task requirements. Relying on activation alone, or a subset of items maintained at the focus of awareness, is a poor mechanism for adaptive decision making or for ensuring skilled performance (Ericsson & Kintsch, 1995).

Following Cowan (2001), one might argue that activation itself, or residence in the focus of awareness, accrues from some kind of context-dependent selection process (i.e., items are “already retrieved”). Under such a scenario, memory records are selected for maintenance, or for residence at the focus of awareness, on the basis of an analysis of the appropriateness of the record in that situation. As I have argued elsewhere (Nairne, 2002b, 2003), it makes sense to conceptualize the continuous record of immediate experience as a pattern of retrieval cues, to be interpreted and selected, rather than as available items “sitting” in some limited-capacity work space. Maintenance activities such as rehearsal are then more appropriately viewed as multiple acts of *recall*, which, in turn, generate additional records for subsequent interpretation.<sup>2</sup> In this conception, however, item recall is not determined by activation per se, but rather by the outcome of a cue-driven retrieval process. We interpret the remnants of immediate experience, selecting responses that are appropriate for the current situation (see Nairne, 1990).

Once we acknowledge that reproduction is cue driven rather than retrieval free, then trace-based concepts such as decay and activation start to lose their

---

<sup>2</sup>It remains an open question as to how many items can be considered part of an “act of recall.” Practically, of course, people can only recall one item at a time.

meaning and power of prediction. Consider decay: It is conceivable that immediate experience leaves behind a record of activity (e.g., a residual pattern of phonological information from language-based processing), and these features, in turn, might decay spontaneously over time. However, the remaining features, by themselves, cannot predict recall; they are merely cues whose predictive value will depend on the discrimination problem facing the organism. One can easily imagine scenarios in which interpreting the activity record will be helped by the loss of features. For example, suppose a certain proportion of trace features is shared by many other possible targets or interpretations (e.g., cue overload). Losing features over time might help solve the discrimination problem in this case, assuming that the shared features are lost and features that uniquely specify a particular target response remain.<sup>3</sup> The fact that memory over the short term can decrease, stay the same, or even improve over time is consistent with this type of reasoning and inconsistent with the simpleminded activation view (see Nairne, 2002b, 2003, for reviews).

If we conceptualize remembering over the short term from a functional perspective, as an active process of response selection based on residual cues, then it makes little adaptive sense to tie reproduction to activation, or forgetting to decay. Some activation of retrieval cues may be necessary for memory, but activation alone is certainly not sufficient for generating an appropriate response. Performance will depend on the particular cues that are active, as well as on the set of possible responses that could be produced. A given activated feature, or set of features, will help or hinder this interpretation process, depending on the context. Stated more generally, the success of remembering cannot be determined by the characteristics of a cue (its strength or activation) or by the characteristics of a single target response, but only through consideration of the cue to target(s) interaction (see Tulving, 1983).

### **The Encoding-Retrieval Match**

Essentially, the same case applies to the proposal that memory, especially long-term memory, depends directly on the similarity, or match, between retrieval cues and traces, as encoded. This is another widely held view, one that has been touted as an important principle by many in the memory field. Toth and Hunt (1999) suggested that the direct relationship between match and retention represents "one of the most important principles ever articulated about memory" (p. 254). Textbooks and review articles regularly proclaim as well, without reservation, that "memory is best when the retrieval environment matches the environment present during encoding" (Haberlandt, 1999, p. 309).

Appeals to similarity, or the cue-target match, once again ignore the functional context of remembering. Just as activation considers only the proper-

---

<sup>3</sup>One can also imagine that the composition of the target set, or the range of possible trace interpretations, might change on a moment-by-moment basis; thus, it would not be possible to predict until the exact moment of recall which features are helpful or not.

ties of a trace, the encoding-retrieval match focuses on the relationship between a cue and one target response and fails to consider the discrimination problem in effect. From a functional perspective, it is more reasonable to assume that cue-target similarity will help target selection in some circumstances and hurt it in others. Match may be generally correlated with retention, but it is unlikely to be the underlying cause of retention. Instead, it is the ability of the retrieval cue to help us discriminate an appropriate response from an inappropriate response that drives retention, not cue-target similarity per se (see Nairne, 2002a, for an extended discussion of this point).

To illustrate, consider the following thought experiment: Subjects are asked to read aloud a list of homophones presented visually on a computer screen (e.g., *pair, pare, pear, pare, pair, etc.*). At test, we ask everyone to write down the homophone that occurred in the third serial position (*pear*). In one condition, we provide no cues, only the retrieval query; in a second condition, we supply an additional cue—the sound of the target word. Note that we have improved the functional cue-target match in this second condition; there are more overlapping features between the retrieval cue and the encoded trace (we will assume that the auditory information is encoded). However, performance is unlikely to improve. All of the items on the list sound alike, so the sound provides no discriminative information about the appropriate response. We may have improved the encoding-retrieval match, but retention in all likelihood will not follow suit.

One can also easily imagine situations in which improving the cue-target match will hurt the retention of a particular target response. Suppose we bolster the encoding-retrieval match by adding cue-target features that are shared by other competitor responses. There may be an increase in the number of overlapping features between the cue and the target, but the discrimination problem itself (picking the correct target response) suffers. Edward DeLosh and I conducted an experiment several years ago to demonstrate this point (DeLosh & Nairne, 1996): Subjects learned to associate nonwords (e.g., *PAVIS*) with three simultaneously presented word cues (e.g., *ROUND, RED, SOFT*). Two of the words were uniquely associated with a nonword (*ROUND* and *RED* were paired only with *PAVIS*); the third cue (*SOFT*) received additional pairings with other nonwords on the list. At test, subjects attempted to recall each nonword in the presence of (a) one unique cue (*ROUND*), (b) two unique cues (*ROUND + RED*), or (c) one unique cue and the shared cue (*ROUND + SOFT*). The results are shown in Figure 9.1.

In terms of the encoding-retrieval match, moving from one unique cue to two unique cues increases the similarity between the encoding and retrieval environments. Each nonword was learned in the presence of three word cues, and providing two of those words at test more closely reinstates the context of original encoding than merely providing one. It is not surprising that performance improved in the two-unique-cue condition compared with the one-unique-cue condition. However, quite different results were obtained in the condition using one unique and one shared cue. Cue-target similarity also improved in this condition, compared with the single-cue condition—in fact, nominally by the same amount as in the two-unique-cue condition—but overall

