Models in search of a brain

BRADLEY C. LOVE
University of Texas, Austin, Texas

AND

TODD M. GURECKIS
Indiana University, Bloomington, Indiana

Mental localization efforts tend to stress the where more than the what. We argue that the proper targets for localization are well-specified cognitive models. We make this case by relating an existing cognitive model of category learning to a learning circuit involving the hippocampus, perirhinal, and prefrontal cortices. Results from groups varying in function along this circuit (e.g., infants, amnesics, and older adults) are successfully simulated by reducing the model's ability to form new clusters in response to surprising events, such as an error in supervised learning or an unfamiliar stimulus in unsupervised learning. Clusters in the model are akin to conjunctive codes that are rooted in an episodic experience (the surprising event) yet can develop to resemble abstract codes as they are updated by subsequent experiences. Thus, the model holds that the line separating episodic and semantic information can become blurred. Dissociations (categorization vs. recognition) are explained in terms of cluster recruitment demands.

Franz remarked in his 1912 essay "New Phrenology" that "the individual parts of the brain do not work independently; they work interdependently, and it is because of the possible functional and anatomical connections that certain types or kinds of mental states are more in evidence than others"(p. 327). To Franz, the allure of localizing mental activities in the brain begot overly simplistic and crude theories of mental processes and brain function.

Localizing mental function need not be problematic. The issue is what to localize. The value of a theory that localizes mental function lies in both the characterization of the mental process and the bridge theory that links this characterization to the brain. Starting with an ill-specified or folk psychological theory of mental function ultimately limits the value of the overall enterprise and invites comparison with Franz's (1912) new phrenology.

In this article, we will argue that well-specified process models of cognitive functions are the appropriate targets for localization. Successful cognitive models, which are quantitatively validated on a broad range of data sets, offer a number of advantages over folk psychological, ad hoc, or traditional psychological theories. In addition to being predictive, behavioral models have mechanisms and dynamics that can be related to brain measures. Although they are not naive accounts of mental function, cognitive models are typically idealized and relatively simple. This clarity provides a good starting point for the localizing of function. Given that debates persist over the basic functions of areas as well studied as the hippocampus (Eichenbaum, 1999; Stark, Bayley, & Squire, 2002), starting simple makes sense.

Here, we will focus on relating a successful process model of human category learning to a learning circuit involving the hippocampus, perirhinal, and prefrontal (PFC) cortices. Category learning offers a good test domain for our approach, since category learning by example is a ubiquitous, flexible, and critical human behavior that is well studied and modeled within cognitive psychology. Furthermore, in recent years, there has been an increasing interest in the cognitive neuroscience of category learning (Ashby & Ell, 2001; Kéri, 2003).

The model we will consider, supervised and unsupervised stratified adaptive incremental network (SUSTAIN; Love, Medin, & Gureckis, 2004), will be applied to human learning data from amnesic patients, infants, young adults, and older adults. Work in neuropsychology, electrophysiology, neuroanatomy, brain imaging, and animal lesion studies will be marshaled to support the mapping between SUSTAIN and structures in the brain. SUSTAIN predicts how degraded function in a learning circuit involving the hippocampus, PFC, and perirhinal cortex affects learning performance for various groups. SUSTAIN relates the degree of preserved function to how readily members of a group can individuate events, as opposed to collapsing experiences together into a common gestalt. To foreshadow, the proposed continuum of function is shown in Figure 1. After introducing SUSTAIN, we will relate aspects of SUSTAIN to a learning circuit involving the PFC, hippocampus, and perirhinal cortex, give evidence for our characterization of this learning circuit, and present supportive simulations. This exercise in relating SUSTAIN

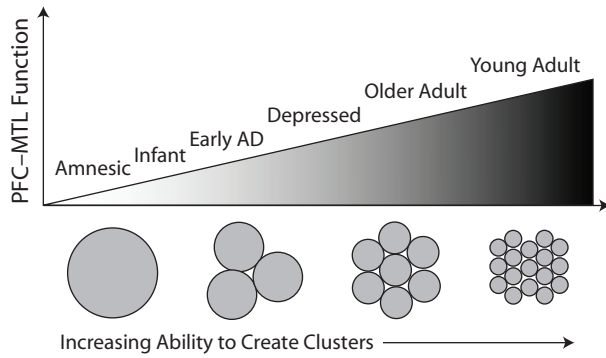B. C. Love, brad_love@mail.utexas.edu

Figure 1. Various groups are ordered by prefrontal cortex–medial temporal lobe (PFC–MTL) function. At one extreme, amnesic patients with hippocampal lesions tend to collapse events into a single gestalt and, therefore, are sensitive primarily to feature or item frequency. At the other extreme, young adults draw finer distinctions and are more sensitive to study context and feature patterns or conjunctions present in their environment. SUSTAIN captures differences along this continuum by varying a parameter related to how readily additional clusters (i.e., bundles of features) are recruited during category learning. AD, Alzheimer's disease.

to particular brain regions suggests a recasting of several dichotomies popular in the field, such as the distinctions between categorization and recognition, recollective and familiarity-driven responding, and episodic and semantic memory.

A Brief Overview of SUSTAIN

The medium for representing category knowledge has been proposed to be rule based (Nosofsky, Palmeri, & McKinley, 1994), exemplar based (Kruschke, 1992; Medin & Schaffer, 1978; Nosofsky, 1986), and prototype based (Rosch & Mervis, 1975; Smith, 2002). SUSTAIN proposes that clusters, which display characteristics of all three of the aforementioned approaches, underlie our category representations (see Love, 2005). A cluster is a bundle of features that captures conjunctive relationships across features (e.g., *wings*, *flies*, and *has feathers* tend to co-occur).

In SUSTAIN, categories are represented by one or more clusters. For example, the category of birds is likely represented by multiple clusters (e.g., song birds, birds of prey, penguins, ostriches, etc.). A cluster can also serve to represent multiple categories. For example, the same cluster can be used to represent information about penguins, birds, and animals. Clusters are linked to categories by association weights that are adjusted by learning rules.

As is shown in Figure 2, SUSTAIN's clusters mediate the relationship between inputs (e.g., stimulus presentation) and output (e.g., category assignment). SUSTAIN begins with one cluster centered on the first training item. Additional clusters are recruited in response to surprising events. In unsupervised learning, a surprising event is exposure to a sufficiently unfamiliar or novel stimulus. In supervised learning, a surprising event is a classification error (e.g., incorrectly predicting that a bat is a bird). SUSTAIN's recruitment scheme implies that surprise drives differentiation of critical stimulus patterns. When

a surprising event does not occur, the current stimulus is assigned to the dominant cluster (i.e., the cluster most activated or similar to the current item), and this dominant cluster moves toward the current stimulus, so that the cluster converges to the centroid or prototype of its members. Thus, in the absence of surprise, events are collapsed together in memory.

As a result, SUSTAIN is a prototype model when each category is represented by one cluster and is an exemplar model when each item is captured by its own cluster. Of course, most cases fall in between these two extremes. In general, the simpler or more regular the category structure, the smaller the number of clusters recruited is (cf. Feldman, 2000). Rule-based influences arise from SUSTAIN's attentional mechanism, which biases it toward solutions that involve a limited number of stimulus attributes.
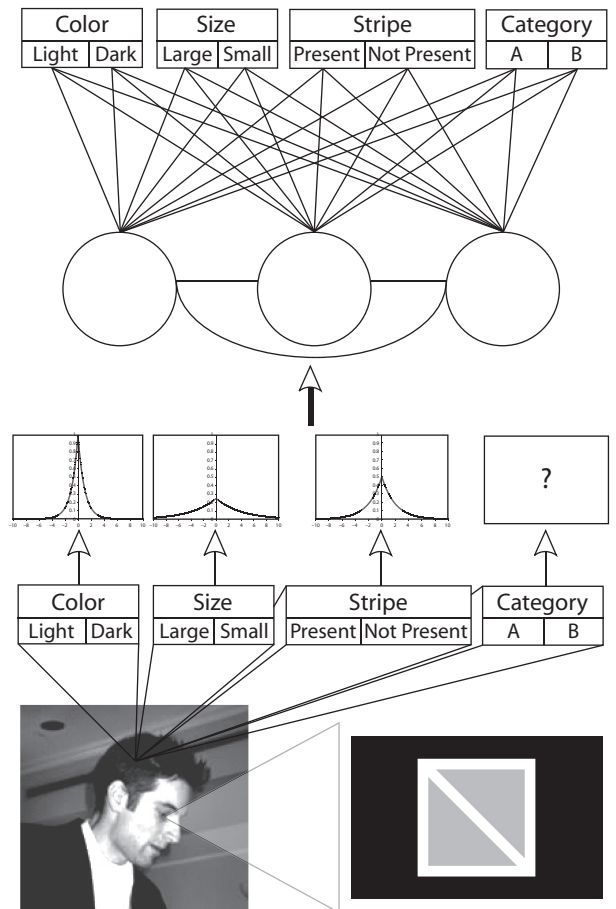

Figure 2. The basic components of the SUSTAIN model. First, the stimulus is encoded along its dimensions (in this case, there are four binary-valued attributes: three perceptual dimensions and the category label). The representational space is contorted (shrunk or stretched along each dimension) by the attentional mechanism to accentuate critical attributes. The clusters (in this case, there are three) compete to respond to the stimulus. The cluster closest to the stimulus in representational space wins (through cluster competition; note the inhibitory connections among the three clusters). The winning cluster predicts the unknown stimulus dimension value (in this case, the category label) by sending a signal to the output units forming the queried dimension.

Although intuitive and simple, SUSTAIN has accounted for an array of challenging data sets spanning a variety of category-learning paradigms, including classification learning (Love & Medin, 1998b), learning at different levels of abstraction (Love & Medin, 1998a), inference learning (Love, Markman, & Yamauchi, 2000), development trends in learning (Gureckis & Love, 2004), unsupervised learning (Gureckis & Love, 2002, 2003), the influence of culture on conceptual organization (Love & Gureckis, 2005), and schematic influences on category learning and recognition memory (Sakamoto & Love, 2004). SUSTAIN's formal description is presented in the Appendix.

SUSTAIN's clustering approach addresses a fundamental challenge facing all learning models: defining what constitutes an event or episode. For example, what constitutes an exemplar in an exemplar-based approach is typically left undefined (see Logan, 1988). To illustrate the problem, consider a learner focusing on a chair while walking across a room. At every moment, he or she is exposed to a slightly different image. The viewpoint is constantly changing and, with it, changes a number of the chair's properties (e.g., the visible features, albedo, etc.). After the learner has walked across the room, is one exemplar stored or are a million? What constitutes an event or experience when one moves outside laboratory-defined learning trials?

SUSTAIN's solution is to collapse all information into the dominant cluster and to recruit a new cluster only in response to a surprising event. In the example above involving the chair, all the information above would be integrated into a single cluster unless something unexpected was encountered. SUSTAIN does not merely replace one problem (defining what an exemplar is) with another (defining what a cluster is) either. SUSTAIN specifies when and how clusters are formed and updated. SUSTAIN's clustering method may prove useful in understanding how humans individuate in general (see Barsalou, Huttenlocher, & Lamberts, 1998). In terms of comparing the performances of various groups (e.g., infants, young adults, older adults, and amnesic patients), SUSTAIN explains how differences in the ability to individuate events lead to different patterns of generalization.

Proposed Mapping Between SUSTAIN and the Brain

Central to SUSTAIN is the ability to form new clusters in response to surprising events. This type of learning is rapid and involves forming new codes or clusters to support subsequent learning. A mature and intact learning circuit involving the hippocampus, PFC, and perirhinal cortex is assumed to underlie this ability. We view the hippocampus as the constructor of new codes or clusters. We propose that the activation of previously recruited clusters is reflected by a familiarity signal generated by structures in the medial temporal lobe (MTL), such as the perirhinal cortex. The PFC is assumed to play a role in directing encoding and retrieval. The PFC plays a critical role in orienting encoding toward surprising events (Corbetta & Shulman, 2002), which dovetails with SUSTAIN's surprise-driven cluster recruitment. In terms of SUSTAIN,

the PFC can be seen as determining when the measure of fit generated by the perirhinal cortex is insufficient. When the current stimulus is judged to be sufficiently surprising, the hippocampus is directed to create a new cluster.

In support of our proposal, the PFC and the perirhinal cortex are interconnected and participate in a circuit that may direct the hippocampus's encoding of surprising events (see Ranganath & Rainer, 2003, for a review). Indeed, lesioning the connection between the PFC and the perirhinal cortex eliminates memory advantages for surprising items (Parker, Wilding, & Akerman, 1998). The PFC monitors surprise by comparing the current stimulus with representations in the perirhinal cortex (which provides a measure of familiarity or fit) and, on the basis of this comparison, directs hippocampal encoding (see Brown & Aggleton, 2001; Ranganath & Knight, 2003). The perirhinal familiarity signal is not bereft of associative or conjunctive information, since it is based on cluster activations, consistent with the position that conjunctive codes are present in regions neighboring the hippocampus (e.g., Stark et al., 2002). Further evidence for the operation of this circuit comes from interactions observed between the PFC and the hippocampus. Suppression of memory for an item leads to increased PFC activation, decreased hippocampal activation, and reduced item memory (M. C. Anderson et al., 2004). Although not as broadly construed, existing learning models have hinted at the learning circuit identified here (Carpenter & Grossberg, 1993; Li, Naveh-Benjamin, & Lindenberger, 2005).

In focusing on the hippocampal aspect of the learning circuit, existing accounts of hippocampal function fit well with our account. Recruiting clusters allows SUSTAIN to form new codes that enable learning of conjunctive or configural information. This process is in the same spirit as configural associative theory, which holds that the hippocampus binds two or more separated representations to create a new unit that enables configural responding (Sutherland, McDonald, Hill, & Rudy, 1989). Similarly, flexible relation theory holds that the hippocampus is specialized for encoding relations among elements (N. J. Cohen & Eichenbaum, 1993). One unique characteristic of our account is that new codes begin as context-sensitive clusters that encode all the attended aspects of the surprising event but, over time, can come to resemble abstract codes as clusters are activated by similar events.

As in Schmajuk and DiCarlo's (1992) model, disabling the hippocampus (i.e., the ability to form multiple clusters in SUSTAIN) prevents rapid learning of novel stimulus configurations. Interpreting the weight update from SUSTAIN's clusters to outputs as cortical learning, SUSTAIN, like Gluck and Myers's (1993) model, assumes that the hippocampus builds the internal codes that support cortical learning. SUSTAIN's rapid recruitment of clusters in response to surprising events parallels Norman and O'Reilly's (2003) hippocampal network, which rapidly acquires distinct episodic traces that bind individual stimulus elements together in memory.

To summarize, we propose that the hippocampus constructs codes, the perirhinal cortex generates a familiarity or fit signal, and the PFC monitors and directs encoding

and retrieval processes. In terms of SUSTAIN, cluster activations relate to the fit signal generated by the perirhinal cortex, with cluster evaluation processes carried out by the PFC. When an event is deemed surprising by the PFC, the hippocampus attempts to construct a new cluster. The role of these areas, as well as their relation to aspects of SUSTAIN, is summarized in Figure 3.

Note that disruption anywhere along the learning circuit can result in the failure to encode a surprising event. For instance, without a functioning hippocampus, an event could be very surprising but not result in a new cluster's being formed. Conversely, with a functioning hippocampus and a damaged PFC, a surprising event that should orient attention for encoding could fail to do so. In modeling terms, the degree of PFC and hippocampal function are captured by separate parameters (see the Appendix for details).

Here, we will focus on populations and tasks in which hippocampal function should be the limiting factor. The fact that behavioral data alone cannot determine where the cognitive bottleneck lies highlights the worth of considering a broader set of measures.

The parameter specifying hippocampal function varies along a continuum, with amnesic patients lacking a hippocampus at one extreme and young normals at the other extreme. The parameter sets SUSTAIN's ability to form clusters that are similar to existing clusters. When the parameter is set low (poor hippocampal function), SUSTAIN can successfully form a new cluster only when the current stimulus is drastically different from any existing cluster. With low functioning, SUSTAIN has trouble establishing new clusters that are somewhat similar to existing clusters and experiences (i.e., events or episodes tend to be undifferentiated). In all the simulations reported here, this parameter is the only aspect of SUSTAIN that varies across populations.

The proposed mapping between SUSTAIN and brain regions is coarse (e.g., no distinction is made between CA1 and CA3 in the hippocampus) and incomplete, since long-term consolidation processes are not considered and other brain areas implicated in category learning are left unmapped. A complete theory of how the brain supports category learning will need to detail the contributions of numerous learning systems. For example, a frontal circuit involving the head of the caudate nucleus, anterior cingulate, and dorsolateral PFC likely supports verbal or rule-based learning (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Posner & Dehaene, 1994). Although it will not be explored fully here, SUSTAIN's attentional system, which selects a small subset of relevant stimulus properties, naturally maps onto this verbal-learning system (Love, 2003). Other learning systems, such as a procedural-learning system (Ashby et al., 1998) and an implicit-learning system (Reber, Stark, & Squire, 1998), also will not be addressed here. We will focus on tasks and groups that stress the role of the PFC–MTL learning circuit in category learning.

By focusing on the hippocampus's role in constructing conjunctive codes, we are not ruling out that other brain areas could carry out similar functions (perhaps with less efficiency and speed). However, we are stating that the hippocampus plays an important role in fulfilling this function. Despite its coarseness and incompleteness, our simple account relating SUSTAIN to the brain will address a diverse set of findings and suggest a number of predictions, which will be discussed in the remainder of the article.

The Nature of Conjunctive Information

Recruiting multiple clusters is necessary for the encoding of conjunctive information. Clusters can be seen as con-
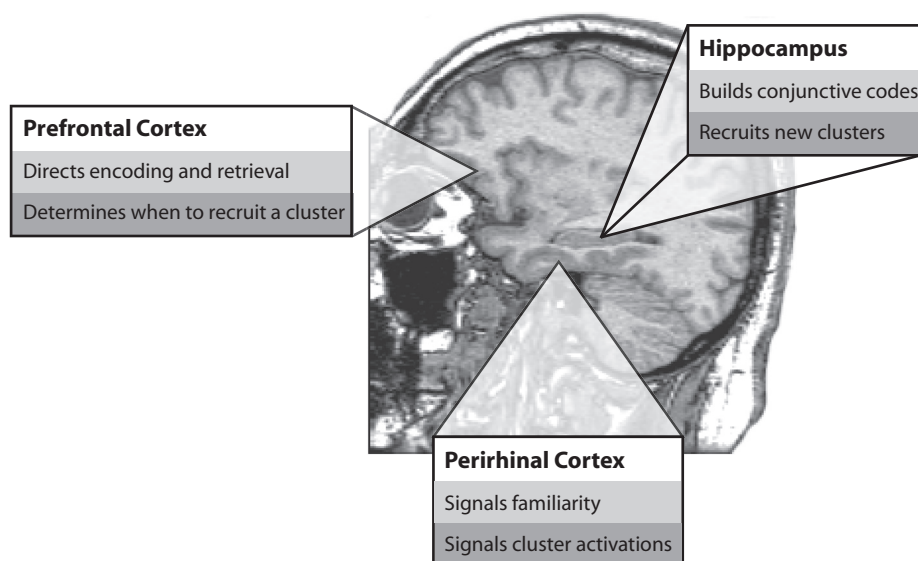

Figure 3. The hippocampus, perirhinal, and prefrontal cortex form a circuit whose operation parallels SUSTAIN's. Each area's functional role in the prefrontal cortex–medial temporal lobe circuit is given, along with its interpretation within SUSTAIN.

junctive codes that bundle together featural information. Clusters capture critical patterns of feature co-occurrence. For example, relationships such as *wooden spoons tend to be large, whereas steel spoons tend to be small*, can be captured by multiple clusters that bundle together related features that define informative subtypes or conjunctions.

How conjunctive a task is and, therefore, how much it relies on the PFC–MTL learning circuit is determined by how many clusters are required for successful performance. Example tasks that will tend to require multiple clusters include memory for item order (a conjunction of item and position), recollection and episodic memory (item or event conjoined with context), list discrimination (a conjunction of item and study list), and memory for item pairs (a conjunction of items). Although SUSTAIN has not yet been applied to all of these tasks, it would require multiple clusters to master such tasks.

In contrast, tasks that are more familiarity driven, such as *know*, as opposed to *remember*, responses in the remember–know paradigm (see Tulving, 1985) and single-item recognition for featurally distinct stimuli, should rely more on existing representations in the perirhinal cortex. SUSTAIN can successfully perform these tasks by utilizing existing clusters or by relying on one recruited cluster, which is sufficient to encode feature frequencies, but not feature co-occurrence patterns.[1]

Although we do claim that the hippocampus is specialized for encoding conjunctive information in the form of clusters, we are not claiming that tasks that rely heavily on conjunctive code formation are qualitatively distinct from those that tend to be less demanding of code construction. For example, although recollective responses tend to require the formation of codes to encode the conjunction of context and item, certain familiarity-based responses may also rely on conjunctive codes. For example, detecting patterns or correlations of features in the environment requires establishing conjunctive codes. In the absence of contextual information, pattern-following items will elicit know responses. Analogously, certain single-item recognition tasks may require some degree of conjunctive encoding (e.g., item feature conjunctions) to successfully differentiate studied and nonstudied items, and amnesic patients should perform worse than normals at such tasks.

Our stance is that there are a number of tasks and methods that have shown dissociations between normals and groups low in hippocampal function and that the true underlying variable, which other variables correlate with, is the number of clusters required for successful performance. Even basic distinctions, such as episodic versus semantic memory, are blurred by our analysis. Reasonably, SUSTAIN predicts that semantic knowledge is rooted in an episode (i.e., a recruited cluster). When a recruited cluster responds to subsequent events and is updated, some of the original episodic features of the cluster will be washed out in the averaging process involved in updating a cluster with new information. Thus, over time, the distinction between episodic and semantic knowledge can become blurred (see McClelland, McNaughton, & O'Reilly, 1995). Likewise, we do not restrict hippocampal

learning to learning involving awareness; rather, we view awareness as an emergent property of hippocampal learning (see Chun & Phelps, 1999; Eichenbaum, 1999).

Corroborating Empirical Data

The next two sections will provide empirical evidence for our account of the PFC–MTL learning circuit, focusing on the role of the hippocampus. The first section will review evidence that the hippocampus is involved in the construction of new codes that support later learning. The second section will review evidence that these codes are conjunctive in nature.

The hippocampus as code builder. Electrophysiological studies of monkeys have shown that new associations for location/scene pairings are encoded in the hippocampus (Wirth et al., 2003). In humans, novel items tend to activate prefrontal areas and the MTL, with particularly robust activations in the hippocampus (Kirchhoff, Wagner, Maril, & Stern, 2000; Knight, 1996; Schacter & Buckner, 1998), and these MTL activations are predictive of subsequent memory (Brewer, Zhao, Desmond, Glover, & Gabrieli, 1998; Grunwald, Lehnertz, Heinze, Helmstaedter, & Elger, 1998; Stern et al., 1996). Ranganath and Rainer (2003) have provided an extensive review of lesion, event-related potential, and imaging studies that demonstrate a novelty response in the PFC and areas of the MTL. The involvement of prefrontal areas in concert with the MTL is consistent with our hypothesis that prefrontal areas are involved in the assessment of surprise and direct encoding of information in the hippocampus.

Time course data from imaging studies also support the idea that the hippocampus plays a prominent role in code construction. Stronger hippocampal activations are often seen early in learning, suggesting that the codes for supporting later learning were initially constructed by the hippocampus (Poldrack et al., 2001; Zeineh, Engel, Thompson, & Bookheimer, 2003). These findings are consistent with the view that codes previously constructed by the hippocampus can support learning, throughout the brain, that does not depend on the hippocampus (Gluck, Oliver, & Myers, 1996). Analogous results have been found in electrophysiological studies in which rabbit eyeblink conditioning has been examined (Sears & Steinmetz, 1990). Consistent with these results, SUSTAIN also tends to recruit clusters early in learning when surprising events are numerous. Subsequent learning utilizes these clusters in the absence of additional cluster recruitment.

Conjunctive codes and the hippocampus. Findings from amnesic patients with hippocampal lesions bolster this theoretical outlook. Amnesic patients tend to be worse at recollection and recall, which requires encoding a conjunction of item and context, than at familiarity-based judgments (Yonelinas, Kroll, Dobbins, Lazzara, & Knight, 1998; Yonelinas et al., 2002). Amnesic patients equated with controls for item recognition remain impaired in list discrimination (Downes, Mayes, MacDonald, & Hunkin, 2002; Mayes et al., 2001), which requires encoding the conjunctions of item and list. In comparison, prefrontal patients are not impaired at this test when equated with controls on item recognition (Mangels, 1997).

Patients with hippocampal damage make more conjunctive errors than do control participants (Kroll, Knight, Metcalfe, Wolf, & Tulving, 1996). In Kroll et al., amnesic patients studied pairings of syllables and were able to recognize studied syllables, but they could not determine whether two syllables had appeared together at study and would false alarm to novel combinations of studied syllables. Similarly, amnesic patients perform well on single-item recognition tasks but show deficits on tasks that depend on encoding relations between items (Holdstock et al., 2000; Vargha-Khadem et al., 1997), such as recognizing the pairing of a picture with a word or a face with a voice. Lesioning studies provide a similar view of hippocampal function. The rat hippocampus supports recognition of spatial arrays of objects, whereas the perirhinal cortex responds to single objects (Wan, Aggleton, & Brown, 1999). This pattern of findings with lesioned rats extends to nonspatial domains (see Eichenbaum, 2000, for a review).

Findings from amnesic patients and lesion studies are corroborated by fMRI studies of healthy normals. Activation of the hippocampus at study predicts recollection, but not item recognition, with the opposite being true for the perirhinal cortex (Davachi, Mitchell, & Wagner, 2003; Ranganath et al., 2004). During retrieval, hippocampal activity increases when accompanied by remember responses, but not when accompanied by know responses (Eldridge, Knowlton, Furmanski, Bookheimer, & Engel, 2000). Recollective memory for scenes is accompanied by hippocampal activation at retrieval, whereas familiarity-based responding is associated with perirhinal activation (Daselaar, Fleck, & Cabeza, 2006; Montaldi, Spencer, Roberts, & Mayes, 2006). Similarly, during recognition, less perirhinal activation is observed for familiar pictures, but more hippocampal activation is observed for successful recognition of picture and name pairings (Gabrieli, Brewer, Desmond, & Glover, 1997).

It is not the case that conjunctive information is simply more difficult to process. One source of evidence arguing against this conclusion is the aforementioned studies in which amnesic patients were equated on item recognition measures but still showed substantial deficits on conjunctive memory tasks. Another source of evidence comes from work showing slower forgetting curves for conjunctive information in healthy normals (Hockley, 1992; Naveh-Benjamin, 2000), which would not be expected if conjunctive information was simply more difficult to encode.

Category Learning and Recognition in Amnesic Patients

As is shown in Figure 1, amnesic patients are characterized by their inability to individuate events (i.e., to recruit clusters in response to surprising events). Knowlton and Squire's (1993) experiments illustrate this point with a category-learning task. Knowlton and Squire found that amnesic patients can categorize, but not recognize, dot pattern stimuli at accuracy levels comparable to those of matched controls.[2] This basic pattern of results has been replicated utilizing stimuli with discrete features (Reed, Squire, Patalano, Smith, & Jonides, 1999).

In Knowlton and Squire's (1993) categorization task, participants viewed 20 low and 20 high distortions of an underlying prototype during the study phase (see Figure 4). The participants were then informed that all of these items belonged to a common category. At test, the participants indicated whether the presented stimulus was a member of the category. The four types of test stimuli included are shown in Figure 4. Half of the test trials consisted of foil patterns that did not conform to the prototype underlying the study items. An example stimulus is shown in Figure 4 under the heading "Random." The other half of the test trials included the presentation of the prototype (which was not actually shown during the study phase), novel low distortions of the prototype, and novel high distortions of the prototype.

In the recognition task, the participants viewed five distinct patterns eight times each. At test, the participants were shown these five items and five foils and indicated whether the stimulus was shown in the study phase. The main results, illustrating a dissociation between categorization and recognition performance for amnesic patients, along with SUSTAIN's fit of these data, are shown in Figure 5. In the categorization test phase, both groups (and SUSTAIN) displayed a generalization gradient that fell as similarity to the prototype decreased (see Figure 6).

Simulations of Knowlton and Squire (1993). Low hippocampal function in SUSTAIN was modeled by reducing the model's ability to recruit a cluster in response to a surprising event when existing clusters were somewhat similar to the surprising stimulus. A parameter regulated the threshold for successfully forming a new cluster in response to a surprising event. Groups low in hippocampal function had a lower setting of this parameter. SUSTAIN simulations of amnesic patients were identical to simulations of the controls, except for the setting of the hippocampal function parameter. Details for all the simulations are provided in the Appendix.
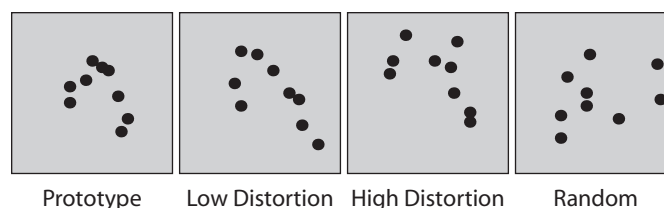

| Prototype | Low Distortion | High Distortion | Random |

Figure 4. Example stimuli from Knowlton and Squire's (1993) categorization task.
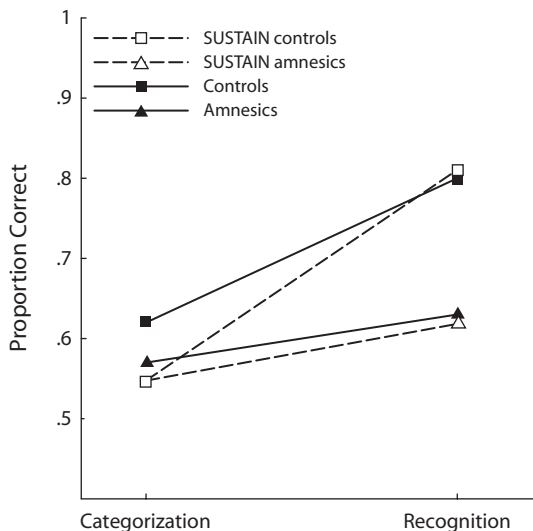
Figure 5. The main results from Knowlton and Squire (1993), along with SUSTAIN's fit.

SUSTAIN's simulations for both groups utilized a single cluster for the category-learning task. All of the study items were sufficiently similar to one another that no surprising events occurred, and therefore, all the items were collapsed into a single cluster. Because both groups had the same internal representation of the study items (i.e., one cluster), SUSTAIN necessarily predicted equivalent performance for the two groups.

In contrast, the simulations for the controls in the recognition task resulted in five clusters being recruited. The five distinct patterns shown in the recognition study phase were sufficiently dissimilar that they were individuated (i.e., each item was surprising when initially presented). In the amnesic simulations of the recognition task, each item was also surprising when initially presented, but because

Figure 6. The generalization gradient for participants from Knowlton and Squire's (1993) categorization test phase are shown, along with SUSTAIN's fit.

of low hippocampal function, clusters were not always recruited in response to these surprising events. Instead of recruiting five clusters, as in the normal simulations, SUSTAIN recruited 2–4 clusters in the amnesic simulations.[3] Collapsing items together within a cluster led to representations that were not well suited for item recognition. Like amnesic patients, the amnesic simulations displayed degraded, but above-chance, recognition performance.

Multiple systems? The theoretical interpretation of SUSTAIN's fits differs from Knowlton and Squire's (1993) interpretation of their data. Following Nosofsky and Zaki (1998), we suggest that recognition and categorization engage common mechanisms (see Love, 2002b, for supportive results with young normals). In fact, as is described fully in the Appendix, recognition and categorization are modeled in an identical fashion. The two tasks used by Knowlton and Squire differ to the degree that they require multiple clusters. If the stimulus sets for recognition and categorization were swapped, we would predict that the pattern of results would be reversed.

In summary, the underlying variable for explaining performance differences between the two populations is not recognition or categorization but reliance on conjunctive codes (i.e., number of clusters required). In the case of Knowlton and Squire's (1993) design, only one cluster is required for successful categorization, and predicted performance is identical for both groups. However, for a more difficult category-learning task that requires multiple clusters (e.g., one that involves multiple categories or category subtypes), SUSTAIN predicts that amnesic patients should show a deficit, relative to controls, consistent with Zaki's (2004) meta-analysis demonstrating such a deficit across a number of category-learning studies in which amnesic patients and controls were compared.

The SUSTAIN simulations bear a resemblance to Nosofsky and Zaki's (1998) modeling of these data with an exemplar model. Nosofsky and Zaki modeled amnesic patients as having lower sensitivity (i.e., broader generalization gradients for each exemplar) than do controls. Effectively, this led to a blurring or averaging of the internal representations for the amnesic simulations. One interpretation of this blurring operation is that it functionally implements a clustering model in which items are collapsed into common clusters, as in SUSTAIN. Although simulation predictions converge, our account does differ from Nosofsky and Zaki's in that SUSTAIN's simulations indicate that amnesic patients should have difficulty encoding items, as evidenced by the smaller number of clusters recruited in the recognition condition, relative to controls. In other words, what counts as an *exemplar* or event differs with changes in PFC–MTL function (see Figure 1). Finally, whereas Nosofsky and Zaki advocated a single-system view of recognition and categorization, we see it as likely that multiple and overlapping systems contribute to both recognition and categorization, although, in this article, we focus on the PFC–MTL learning circuit.
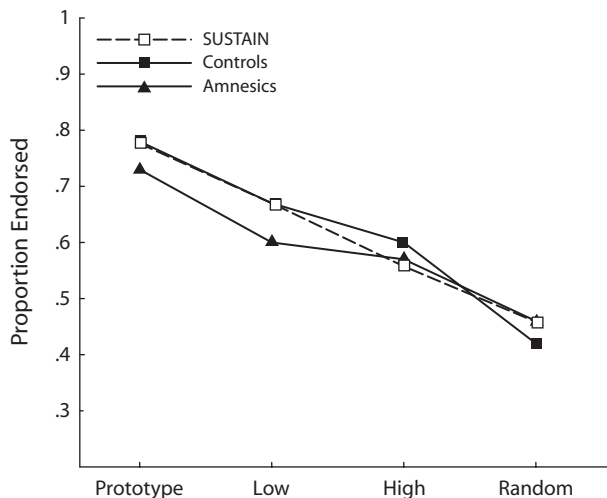
Infant Category Learning

The human brain undergoes considerable development postbirth. In fact, important structural changes continue

up to adulthood. As a first approximation, the pattern of development follows a *back-to-front* trajectory, starting with basic visual areas and progressing toward frontal areas (Johnson, 2003), as evidenced by synaptic growth and glucose uptake (Chugani, Phelps, & Mazziotta, 1987). The hippocampus is unusual in that it undergoes functional neurogenesis throughout adulthood (Eriksson et al., 1998; van Praag et al., 2002), and this growth may be exaggerated in individuals who frequently engage in stimulating behaviors (Maguire et al., 2000; van Praag, Christie, Sejnowski, & Gage, 1999). These new neurons participate in the formation of new hippocampal-dependent memories (Shors et al., 2001) and may naturally map to the cluster formation process in SUSTAIN.

Considerable developmental changes occur in the hippocampus during the 1st year of life. Cell differentiation surges in the hippocampus between 7 and 10 months (Seress & Mrzljak, 1992) and continues into the 2nd year (Merzenich & Sameshima, 1993). Connectivity between the hippocampus and other areas increases into the 2nd year (Benes, 1994). Hippocampal volume reaches adult level during the second half of the 1st year of life (Kretschmann, Kammradt, Krauthausen, Sauer, & Wingert, 1986).

Given these developmental changes, one reasonable hypothesis is that infants' PFC–MTL learning circuit will not be fully functional. In terms of the continuum shown in Figure 1, young infants (4-month-olds) are assumed to have function closer to that of amnesic patients than to that of young adults. Thus, infants should show difficulty in conjunctive tasks that require multiple clusters. Indeed, the ability to encode and, after a delay, retrieve conjunctions of objects and actions begins to emerge at 9 months of age (Carver & Bauer, 2001; Diamond, Churchland, Cruess, & Kirkham, 1999; Nelson, 1995). Toward the end of the 1st year of life, memory capabilities emerge that are more context sensitive or recollective in nature, as opposed to solely familiarity driven (see de Haan, Mishkin, Baldeweg, & Vargha-Khadem, 2006, for a review).

Simulations of Younger and Cohen (1986). Younger and Cohen (1986) conducted a series of experiments well suited for assessing infants' abilities to form conjunctive codes (i.e., multiple clusters). In their experiments, infants were habituated to a set of sequentially presented visual stimuli depicting imaginary animals. The stimuli varied on three binary-valued attributes (e.g., the type of body for an animal was elephant- or giraffe-like). Importantly, two stimulus attributes correlated perfectly across the study items. After being habituated, infants were exposed to test items that either followed the correlated (i.e., conjunctive) pattern of the habituation items or violated it. If infants formed a conjunctive code at study (i.e., during habituation), they should find the item inconsistent with the conjunctive code novel and, therefore, look at it longer. In contrast, if infants failed to encode the attribute relation but, instead, encoded only attribute value (i.e., feature) frequencies, both the consistent and the inconsistent test items should be equally interesting to the infants and should yield equal looking times.

The design of Younger and Cohen's (1986) Experiment 2 is shown in Table 1, and the basic findings, along

Table 1
The Abstract Structure of Younger and Cohen's (1986)
Experiment 2

Study	Test
1 1 1	2 2 2 (consistent)
1 1 2	2 1 2 (inconsistent)
2 2 1	
2 2 2	

Note—The first two attributes of the study items correlate perfectly. After being habituated for 12 trials during study, participants viewed the consistent test item (which obeys the studied correlation) and the inconsistent test item (which violates the studied correlation).

with SUSTAIN's fit, are shown in Figure 7.[4] Four-month-old infants' looking times for the consistent and the inconsistent items were equal, whereas 10-month-old infants devoted more time to the inconsistent item than to the consistent item.[5] SUSTAIN displayed this same pattern. As in the previous simulations comparing amnesic patients and controls, the group with lower PFC–MTL function (in this case, the 4-month-olds) were modeled with a lower setting of the hippocampal function parameter.

Figure 8 shows the spatial configuration of SUSTAIN's clusters, relative to study and test items. In the simulation for 4-month-olds (shown in the left plot of Figure 8), SUSTAIN recruited a single cluster. The single cluster represented the average of the four study items (i.e., the feature frequencies). This single cluster was located in the center of the space and was equidistant from both the consistent and the inconsistent test items. Given this configuration, SUSTAIN predicted that both of the test items would be equally familiar.

In the simulation for 10-month-olds (shown in the right plot of Figure 8), SUSTAIN recruited two clusters. Each of these two clusters represented the average of two of the four study stimuli. One cluster represented the average of stimuli 1 1 2 and 1 1 1 (located at 1 1 1.5), whereas the other represented the average of stimuli 2 2 2 and 2 2 1 (located
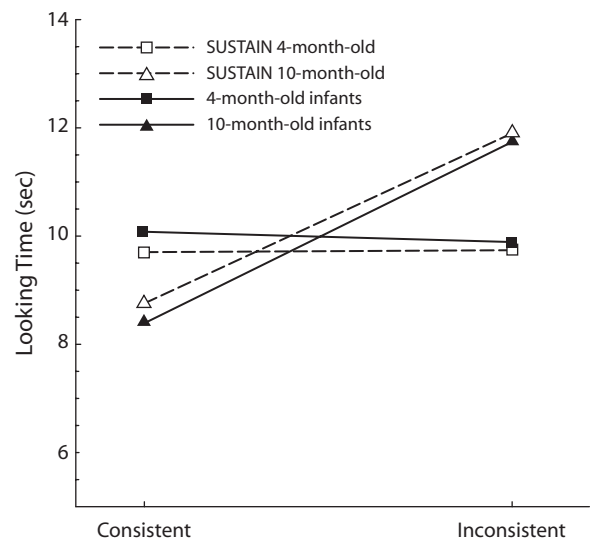

Figure 7. Looking times for 4-month-old and 10-month-old infants in Younger and Cohen's (1986) Experiment 2, along with SUSTAIN's fit.
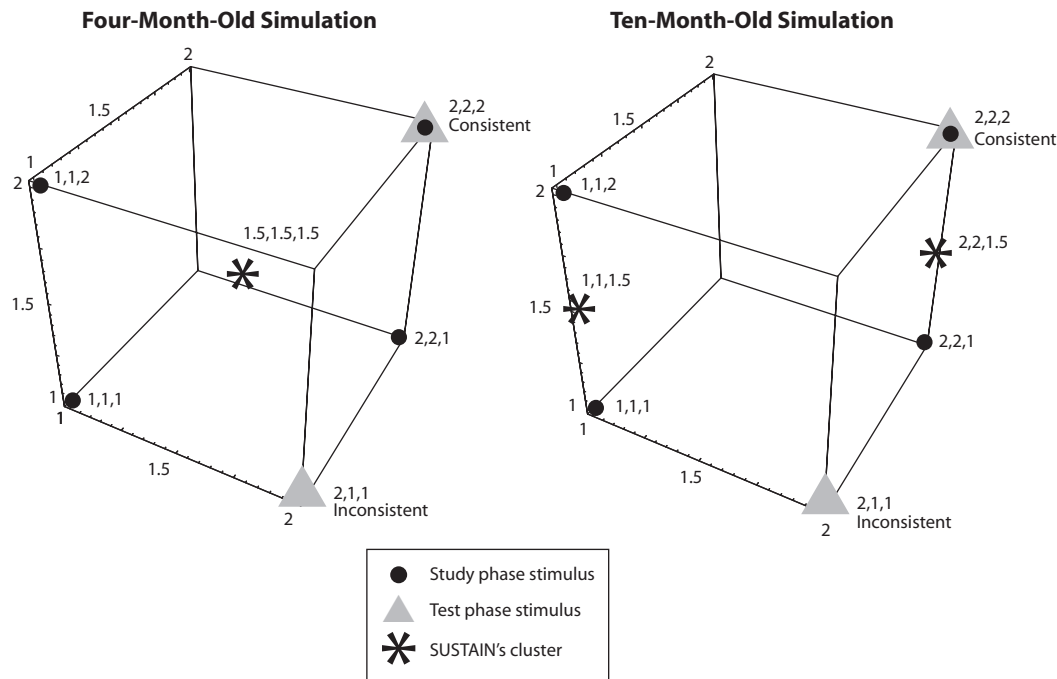
Four-Month-Old Simulation

Ten-Month-Old Simulation

Figure 8. A geometric rendering of the stimulus structure described in Table 1, along with SUSTAIN's cluster-ing solutions. The simulation of 10-month-olds in Younger and Cohen's (1986) Experiment 2 recruited two clus-ters that captured the conjunctive pattern across the first two stimulus attributes, whereas in the simulation of 4-month-olds, the single-cluster representation was sensitive only to attribute value (i.e., feature) frequencies.

at 2 2 1.5). In this case, the inconsistent test stimulus was farther from the nearest cluster than was the consistent test stimulus. This effect was magnified by SUSTAIN's shift of attention to the two correlation-relevant attributes. These two clusters effectively encoded the conjunctive relation-ship between the first two attributes of the study items.

These simulations closely parallel the simulations of Knowlton and Squire's (1993) data. Both young infants and amnesic patients are sensitive to feature frequency, but not to feature relations. Perhaps an even closer parallel to Younger and Cohen's (1986) experiments is Save, Poucet, Foreman, and Buhot's (1992) experiments, in which rats were habituated in an environment consisting of multiple objects. After familiarizing themselves with the environ-ment, the rats' rate of exploration declined, much as the infants in Younger and Cohen's study habituated to the stimuli during the study phase. Once the rats had been habituated, the experimenter altered the configuration of objects. The rats with an intact hippocampus once again found their environment engaging and increased their rate of exploration, which is analogous to the 10-month-olds' reaction to the inconsistent test item. However, the rats who had their hippocampus lesioned did not increase their rate of exploration, which is analogous to the 4-month-olds' lack of differentiation between the consistent and the inconsistent test items.

Aging and Category Learning

Aging does not affect the brain uniformly. Imaging and neuroanatomical studies have revealed greater shrinkage in the hippocampus and PFC than in other areas, such as the parietal and occipital cortex (Flood & Coleman, 1988; Raz, 2000). These declines should impact our proposed PFC–MTL learning circuit, and therefore, we predict that older individuals will show deficits in tasks that require constructing conjunctive codes (see Davidson & Glisky, 2002; Li et al., 2005).

The hippocampus is particularly vulnerable in the aging process. The hippocampus has the highest concentration of glucocorticoid (i.e., cortisol) receptors in the central ner-vous system (McEwen, Weiss, & Schwartz, 1968). Cortisol released in response to agitating or stressful episodes (e.g., illness, trauma, surgery, temperature extremes, depressive states, or anxious moods) leads to hippocampal shrinkage and loss of function. One possibility is that the aging ef-fects seen in the hippocampus result simply from the ac-cumulation of negative events over an individual's life.

Supporting this notion, an MRI study of humans has shown that hippocampal atrophy is the anatomical cor-relate of delayed recall performance in older adults (Golomb et al., 1994). Humans with prolonged elevated levels of cortisol showed reduced hippocampal volume and deficits in hippocampal-mediated memory tasks (Lu-pien et al., 1998). Experimental manipulations of stress level with rats led to predicted increases in cortisol levels, hippocampal atrophy, and memory impairments (Land-field, Baskin, & Pitler, 1981; Sapolsky, Krey, & McEwen, 1986). Underscoring the importance of cortisol in cog-nitive aging, recent work suggests that elevated cortisol levels may also affect frontal function in humans (Young,

Sahakian, Robbins, & Cowen, 1999). Paralleling the negative effects of excessive cortisol, older adults showed reduced activations, relative to young adults, in the PFC and hippocampus during encoding (Grady et al., 1995), which maps onto reduced cluster recruitment (i.e., conjunctive code formation) in SUSTAIN.

Given these assaults on the PFC–MTL learning circuit, we predict that older adults should be impaired at conjunctive-learning tasks. Indeed, item memory in older individuals is preserved to a greater extent than is memory for the conjunction of item and context (see Spencer & Raz, 1995, for a meta-analysis). Recollection (a conjunction of item and context) and source memory are impaired in older individuals, whereas some studies have shown no impairment in judgments of familiarity (Dywan & Jacoby, 1990; Hay & Jacoby, 1999; Titov & Knight, 1997). For example, older adults have trouble determining at test which modality an item was presented in at study (McIntyre & Craik, 1987).

These deficits hold for other conjunctive tasks that involve forming arbitrary associations or conjunctions. Older adults are particularly bad at forming conjunctive codes, such as linking or associating two pictures (Chalfonte & Johnson, 1996; Naveh-Benjamin, Hussain, Guez, & Bar-On, 2003). Age-related declines in hippocampal activation, with accompanying behavioral deficits, have been observed in conjunctive memory tasks (Mitchell, Johnson, Raye, & D'Esposito, 2000). Interestingly, such conjunctive tasks are not as difficult when they rely on preexisting associations or codes (Naveh-Benjamin, 2000; Hess, Pullen, & McGee, 1996). In such cases, SUSTAIN predicts that a new cluster is not needed, so hippocampal involvement should be minimal.

In the domain of category learning, Hess (1982) has demonstrated that older adults can abstract prototypes as effectively as can younger adults. Later work incorporating a distraction task showed prototype organization following study for both younger and older adults, with the younger participants having a slight advantage. However, the advantage was much more pronounced for a disjunctive concept-learning task (Hess & Slaughter, 1986) that, according to SUSTAIN, requires forming multiple clusters. Paralleling the amnesic findings, these aging findings

suggest that differences between older and younger adults are magnified for tasks that require forming conjunctive codes. Consistent with this view, Ashby, Noble, Filoteo, Waldron, and Ell (2003) found that contrastive category-learning tasks involving irregularly structured categories (requiring multiple clusters to be mastered) were more difficult for older adults, whereas differences between younger and older adults were minimal for easily describable structures (also see Filoteo & Maddox, 2004).

Simulating the effects of aging on category learning. Ideally, a study in which younger and older adults were examined would utilize a single task in order to compare reliance on conjunctive codes both between and within groups. One such study (Love, 2002a; Love & Gureckis, 2004; Love, Gureckis, & Worchel, 2007) was simulated with SUSTAIN. In Love et al., younger and older adults were trained by supervised classification learning on the two contrastive categories shown in Table 2. As can be seen in Table 2, there was an imperfect rule on the first attribute of each category. For instance, if the first attribute was randomly assigned to size for a participant and 1 indicated a small object and 2 a large object, all the items in Category A (except the last item) would be small. To further cue this imperfect rule during the study phase, a hint was provided at the bottom of the screen (e.g., "A: size is large; B: size is small"). Cuing the rule increased the likelihood that the deficits observed in older adults are attributable to the PFC–MTL learning system, as opposed to working memory or hypothesis-testing learning systems. Participants completed 80 study trials and then completed a test phase consisting of the eight studied items and eight novel items that contained the same features as the studied items rearranged. In the test phase, the participants indicated the category membership of the stimulus, as in the study phase, but the hint indicating the imperfect rule and corrective feedback were not provided.

The main prediction was that relative to younger adults, older adults should have greater difficulty learning the exception items than the rule-following items. This prediction was based on modeling the older adults by reducing SUSTAIN's ability to form new clusters that were similar to existing clusters (as in previous simulations of amnesic patients and 4-month-old infants). SUSTAIN predicted that exception items could be mastered only by forming separate clusters to encode these items. Because exception items would be fairly similar to existing clusters that captured rule-following items from the opposing category, SUSTAIN predicted that these items should be especially difficult for older adults to master.

These predictions held. The main results from the study phase and SUSTAIN's fit are shown in Figure 9. Older adults performed equivalently to younger adults on rule-following items but showed a large deficit on the exception items. Within the older adult group, the difficulty with exception items increased with age, whereas performance actually increased with age for rule-following items. As can be seen in Figure 10, SUSTAIN created separate clusters for each exception in the younger adult simulations, allowing it to eventually master these items. In contrast, SUSTAIN assigned the exception items to clusters that

Table 2
The Logical Structure of the Stimulus Set for the Study and Test Phases of the Aging Study

Category A	Category B
Study Phase	
1 1 1 2	2 1 1 2
1 1 2 1	2 1 2 1
1 2 1 1	2 2 1 1
2 2 2 2	1 2 2 2
Test Phase	
1 1 1 1	2 1 1 1
1 1 2 2	2 1 2 2
1 2 1 2	2 2 1 2
1 2 2 1	2 2 2 1

Note—The first attribute specifies an imperfect rule, with the fourth item in each category violating the rule.
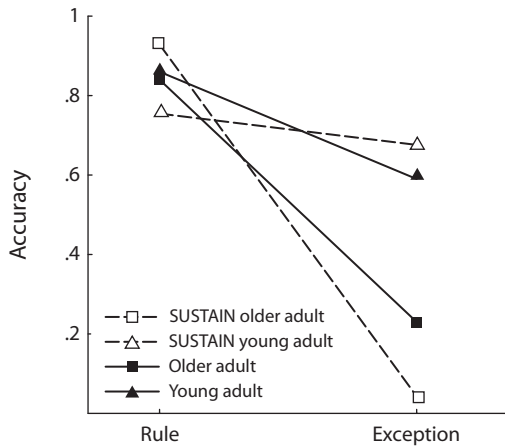
Figure 9. Study phase accuracy for the study phase of the aging study, along with SUSTAIN's fit.

largely captured the rule-following items from the *opposing* category for the older adult simulations. For these simulations, SUSTAIN failed to individuate the exception items and treated these items as if they provided support for the discriminative rule. Rather than simulating individual differences within groups, we chose to simulate idealized younger and older adults. Thus, SUSTAIN overstated the interaction between group and item type shown in Figure 9.

SUSTAIN clusterings for the older adult simulations predicted that older adults should have more abstract rule representations than would younger adults. As is shown in the lower panel of Figure 10, the clusters for the older adult simulations stressed the rule-relevant stimulus dimension and did not preserve item-specific information. This prediction held. Older adults applied the imperfect rule in the transfer phase as often to rule-following items that appeared in the study phase as they did to novel items. In contrast, younger adults' rule application was influenced more by similarity to exemplars seen in the study phase (cf. Allen & Brooks, 1991), as is suggested by SUSTAIN's clusterings shown in Figure 10.

These results distinguish between our approach to capturing reduced function and Nosofsky and Zaki's (1998) approach.[6] Extrapolating Nosofsky and Zaki's amnesic simulations to the aging data, the older adult model would

perfectly store every example in an exemplar memory system, but retrieval operations would blur these exemplars, creating confusions between exemplars. This blurring operation predicts the main result (i.e., older adults' deficit centers on exception items), but it cannot predict the observed stronger rule-based responding with increasing age. Much as Nosofsky and Zaki's amnesic simulations predict lower performance for amnesic patients in both recognition and categorization, with a greater deficit in recognition, analogous simulations of the aging data would predict deficits for both rule and exception items, with a greater deficit in exception performance.

General Discussion

The what of localization is at least as important as the where. For example, the shift from viewing the brain, as opposed to the heart, as the seat of cognition was perhaps the greatest advance in localizing mental function, yet it had little impact, in its own right, on how we understood human behavior. Although our understanding of cognition has advanced beyond that of the early Greeks, we still run the danger of amassing a list of brain areas associated with certain tasks in the absence of linking theories. In this article, we argued that localizing cognitive models (that simulate interesting behaviors) are our best bet for directing and understanding empirical research. At some level, every researcher is theoretically driven and relies on a model of how cognition works, even if that model is not explicitly acknowledged. We argue that the best model to use is one that is well specified, relatively simple, and verified empirically—that is, an existing cognitive model. Successful cognitive models best suited for this task are ones that stress processing and representation, as opposed to models that are formulated at a more abstract level, such as Bayesian approaches. Models that stress simulation form a diverse group, ranging from connectionist models (e.g., SUSTAIN) to production systems, such as ACT–R (J. R. Anderson, Qin, Stenger, & Carter, 2004).

In this article, we applied an existing model of category learning from examples to findings from numerous populations, including amnesic patients with hippocampal lesions, infants, young adults, and older adults. The mapping between the SUSTAIN model and the PFC, hippocampus, and perirhinal cortex was simple and incomplete, yet it was sufficiently powerful to place findings from numerous subfields into a common theoretical framework.

Our account holds that a healthy PFC–MTL circuit is necessary to encode conjunctive information in the form of clusters, which encode relations among stimulus features. New clusters are recruited in response to surprising events, such as an unfamiliar stimulus in unsupervised learning or a prediction error in supervised learning. Thus, new conjunctive codes begin as episodic traces.

Individuals with low PFC–MTL function will fail to establish a new cluster when an existing cluster is somewhat similar to the surprising stimulus. In such cases, the surprising stimulus is assimilated into an existing cluster. Thus, groups varying in hippocampal function will differ in the granularity in which events are individuated (see Figure 1), with low function groups collapsing numerous experiences
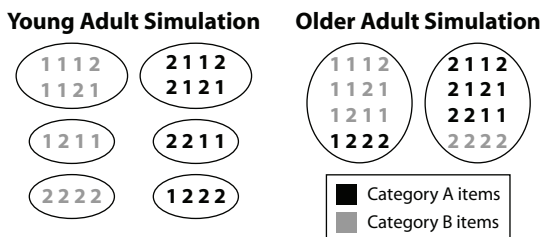

Figure 10. Typical SUSTAIN clustering for the aging study. Rule-violating items were stored separately for the young adult simulation, whereas these items were stored with rule-following items from the contrasting category for the older adult simulation.

together into a single representation (i.e., cluster). These low-function groups will show the greatest deficits in tasks that require establishing numerous conjunctive codes.

In SUSTAIN, these differences between groups are captured by a parameter related to hippocampal function. This parameter governs the ability of the model to form clusters in response to surprising events in the presence of somewhat similar existing clusters (see the Appendix for details). Although the focus is on the hippocampus, other areas of the brain are critical to category learning. Activations of existing clusters are reflected by a familiarity or fit signal generated by the perirhinal cortex. The PFC monitors this signal to determine when a stimulus is sufficiently surprising to warrant the creation of a new cluster. Figure 3 summarizes the mapping between SUSTAIN and these regions.

Although we have argued that cognitive models offer a tool for determining functional localization, accepted views on functional localization may offer an even more powerful tool for selecting among cognitive models that successfully address behavioral findings but make conflicting assumptions about representation and processing. In cases in which these conflicting assumptions suggest different accounts of localization, cognitive neuroscience data can play a key role in selecting the appropriate cognitive model. In the case of our clustering account and Nosofsky and Zaki's (1998) exemplar approach, we believe the mapping between SUSTAIN and the PFC–MTL learning system is more satisfying than possible alignments we can imagine between the exemplar approach and the hippocampus. As has been discussed, Nosofsky and Zaki assumed that all exemplars are stored but are stored in a blurry fashion for populations with reduced function, so that retrieval operations coactivate numerous exemplars in memory. Veridically storing all exemplars (albeit in a fashion that promotes generalization across exemplars) seems at odds with modeling amnesic populations that lack the proposed machinery for storing exemplars. In contrast, our clustering account proposes an interplay among brain areas to determine when storage operations should occur (i.e., in response to surprising events), with the degree of preserved hippocampal function determining when desired storage operations will be successful.

Filling in the continuum. One avenue for future investigation is considering additional populations that suffer from low functioning along the proposed PFC–MTL circuit, such as those suffering from early Alzheimer's disease. Although the damage caused by Alzheimer's disease affects numerous parts of the brain, including frontal areas, early cell loss is concentrated in the hippocampus formation (Coleman & Flood, 1987; Convit et al., 1995; Golomb et al., 1994). These patients exhibit the kinds of learning deficits we would predict (e.g., Faust, Balota, & Spieler, 2001; Zaki, Nosofsky, Jessup, & Unverzagt, 2003). As has been predicted, these patients do not show deficits, relative to age-matched controls, on nonconjunctive categorization tasks that rely on forming only one cluster (Bozoki, Grossman, & Smith, 2006). Finally, those suffering from chronic depression and exhibiting hippocampal atrophy related to excess cortisol levels show the kind of behavioral deficits we would predict (Butters et al., 2000; Golinkoff & Sweeney, 1989).

In all of our simulations and discussions, we have focused on one problem in cluster formation—namely, the lack of it. There are many other possible disorders to explore as a cluster formation process. One possibility is that some groups, such as autistics, recruit new clusters too aggressively and this is the cause for their poor generalization beyond the training set (see I. L. Cohen, 1994; Klinger & Dawson, 2001). In effect, autistic patients may be to the right of young normals along the continuum shown in Figure 1.

Limits and future challenges. Although our account of PFC–MTL function is applicable to a wide range of populations, it is not applicable to all populations and tasks. For example, our account is silent on Parkinson's patients with damage concentrated in striatal dopamine-driven learning systems. Accordingly, our account does not make predictions about behavioral manipulations intended to disrupt procedural learning mediated by the striatum (see Ashby & Maddox, 2005, for a review of relevant manipulations). Finally, our account is not applicable to relevant populations, such as hippocampal amnesic patients, when the behavioral task primarily taps learning systems other than the PFC–MTL circuit. For example, simple rule-based tasks, such as the Wisconsin card sorting task (Heaton, 1981), are likely accomplished through working memory systems that can be preserved in patients with hippocampal damage (Ashby & Maddox, 2005). One important direction for future work will be the development of models that detail the interactions of numerous learning systems.

Beginning with a cognitive model has a number of advantages, but it is just a starting point. The next step is to use brain measures to modify our proposed learning algorithms. Given that the hippocampus is central to our explanation, one critical process that needs to be included in future modeling is consolidation. The importance of consolidation to memory performance has been underappreciated by cognitive psychologists (Wixted, 2005). Another example in which brain measures could influence model development is in work in novelty detection. As has been explored in this article, specifying when an event is surprising and results in cluster recruitment is central to our modeling efforts.

REFERENCES

ALLEN, S. W., & BROOKS, L. R. (1991). Specializing the operation of an explicit rule. *Journal of Experimental Psychology: General*, **120**, 3-19.

ANDERSON, J. R., QIN, Y., STENGER, V. A., & CARTER, C. S. (2004). The relationship of three cortical regions to an information-processing model. *Journal of Cognitive Neuroscience*, **16**, 637-653.

ANDERSON, M. C., OCHSNER, K. N., KUHL, B., COOPER, J., ROBERTSON, E., GABRIELI, S. W., ET AL. (2004). Neural systems underlying the suppression of unwanted memories. *Science*, **303**, 232-235.

ASHBY, F. G., ALFONSO-REESE, L. A., TURKEN, A. U., & WALDRON, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, **105**, 442-481.

ASHBY, F. G., & ELL, S. W. (2001). The neurobiology of category learning. *Trends in Cognitive Sciences*, **5**, 204-210.

ASHBY, F. G., & MADDOX, W. T. (2005). Human category learning. *Annual Review of Psychology*, **56**, 149-178.

ASHBY, F. G., NOBLE, S., FILOTEO, J. V., WALDRON, E. M., & ELL, S. W. (2003). Category learning deficits in Parkinson's disease. *Neuropsychology*, **17**, 115-124.

BARSALOU, L. W., HUTTENLOCHER, J., & LAMBERTS, K. (1998). Basing categorization on individuals and events. *Cognitive Psychology*, **36**, 203-272.

BENES, F. M. (1994). Development of the corticolimbic system. In G. Dawson & K. W. Fisher (Eds.), *Human behavior and the developing brain* (pp. 176-206). New York: Guilford.

BOZOKI, A., GROSSMAN, M., & SMITH, E. E. (2006). Can patients with Alzheimer's disease learn a category implicitly? *Neuropsychologia*, **44**, 816-827.

BREWER, J. B., ZHAO, Z., DESMOND, J. E., GLOVER, G. H., & GABRIELI, J. D. E. (1998). Making memories: Brain activity that predicts how well visual experience will be remembered. *Science*, **281**, 1185-1187.

BROWN, M. W., & AGGLETON, J. P. (2001). Recognition memory: What are the roles of the perirhinal cortex and hippocampus? *Nature Reviews Neuroscience*, **2**, 51-61.

BUTTERS, M. A., BECKER, J. T., NEBES, R. D., ZMUDA, M. D., MULSANT, B. H., POLLOCK, B. G., & REYNOLDS, C. F. (2000). Changes in cognitive functioning following treatment of late-life depression. *American Journal of Psychiatry*, **157**, 1949-1954.

CARPENTER, G. A., & GROSSBERG, S. (1993). Normal and amnesic learning, recognition and memory by a neural model of cortico-hippocampal interactions. *Trends in Neurosciences*, **16**, 131-137.

CARVER, L. J., & BAUER, P. J. (2001). The dawning of a past: The emergence of long-term explicit memory in infancy. *Journal of Experimental Psychology: General*, **130**, 726-745.

CHALFONTE, B. L., & JOHNSON, M. K. (1996). Feature memory and binding in young and older adults. *Memory & Cognition*, **24**, 403-416.

CHUGANI, H. T., PHELPS, M. E., & MAZZIOTTA, J. C. (1987). Positron emission tomography study of human brain functional development. *Annals of Neurology*, **22**, 487-497.

CHUN, M. M., & PHELPS, E. A. (1999). Memory deficits for implicit contextual information in amnesic subjects with hippocampal damage. *Nature Neuroscience*, **2**, 844-847.

COHEN, I. L. (1994). An artificial neural network analogue of learning in autism. *Biological Psychiatry*, **36**, 5-20.

COHEN, N. J., & EICHENBAUM, H. (1993). *Memory, amnesia, and the hippocampal system*. Cambridge, MA: MIT Press.

COLEMAN, P. D., & FLOOD, D. G. (1987). Neuron numbers and dendritic extent in normal aging and Alzheimer's disease. *Neurobiology of Aging*, **8**, 521-545.

CONVIT, A., DE LEON, M. J., TARSHISH, C., DE SANTI, S., KLUGER, A., RUSINEK, H., & GEORGE, A. (1995). Hippocampal volume losses in minimally impaired elderly. *Lancet*, **345**, 266.

CORBETTA, M., & SHULMAN, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, **3**, 201-215.

DASELAAR, S. M., FLECK, M. S., & CABEZA, R. (2006). Triple dissociation in the medial temporal lobes: Recollection, familiarity, and novelty. *Journal of Neurophysiology*, **96**, 1902-1911.

DAVACHI, L., MITCHELL, J. P., & WAGNER, A. D. (2003). Multiple routes to memory: Distinct medial temporal lobe processes build item and source memories. *Proceedings of the National Academy of Sciences*, **100**, 2157-2162.

DAVIDSON, P. S. R., & GLISKY, E. L. (2002). Neuropsychological correlates of recollection and familiarity in normal aging. *Cognitive, Affective, & Behavioral Neuroscience*, **2**, 174-186.

DE HAAN, M., MISHKIN, M., BALDEWEG, T., & VARGHA-KHADEM, F. (2006). Human memory development and its dysfunction after early hippocampal injury. *Trends in Neurosciences*, **29**, 374-379.

DIAMOND, A., CHURCHLAND, A., CRUESS, L., & KIRKHAM, N. Z. (1999). Early developments in the ability to understand the relation between stimulus and reward. *Developmental Psychology*, **35**, 1507-1517.

DOWNES, J. J., MAYES, A. R., MACDONALD, C., & HUNKIN, N. M. (2002). Temporal order memory in patients with Korsakoff's syndrome and medial temporal amnesia. *Neuropsychologia*, **40**, 853-861.

DYWAN, J., & JACOBY, L. (1990). Effects of aging on source monitoring: Differences in susceptibility to false fame. *Psychology & Aging*, **5**, 379-387.

EICHENBAUM, H. (1999). Conscious awareness, memory, and the hippocampus. *Nature Neuroscience*, **2**, 775-776.

EICHENBAUM, H. (2000). A cortical-hippocampal system for declarative memory. *Nature Reviews Neuroscience*, **1**, 41-50.

ELDRIDGE, L. L., KNOWLTON, B. J., FURMANSKI, C. S., BOOKHEIMER, S. Y., & ENGEL, S. A. (2000). Remembering episodes: A selective role for the hippocampus during retrieval. *Nature Neuroscience*, **3**, 1149-1152.

ERIKSSON, P. S., PERFILIEVA, E., BJÖRK-ERIKSSON, T., ALBORN, A.-M., NORDBORG, C., PETERSON, D. A., & GAGE, F. H. (1998). Neurogenesis in the adult human hippocampus. *Nature Medicine*, **4**, 1313-1317.

FANSELOW, M. S. (1999). Learning theory and neuropsychology: Configuring their disparate elements in the hippocampus. *Journal of Experimental Psychology: Animal Behavior Processes*, **25**, 275-283.

FAUST, M. E., BALOTA, D. A., & SPIELER, D. H. (2001). Building episodic connections: Changes in episodic priming with age and dementia. *Neuropsychology*, **15**, 626-637.

FELDMAN, J. (2000). Minimization of Boolean complexity in human concept learning. *Nature*, **407**, 630-633.

FILOTEO, J. V., & MADDOX, W. T. (2004). A quantitative model-based approach to examining aging effects on information-integration category learning. *Psychology & Aging*, **19**, 171-182.

FLOOD, D. G., & COLEMAN, P. D. (1988). Neuron numbers and sizes in aging brains: Comparisons of humans, monkeys, and rodent data. *Neurobiology of Aging*, **9**, 453-463.

FRANZ, S. I. (1912). New phrenology. *Science*, **35**, 321-328.

GABRIELI, J. D. E., BREWER, J. B., DESMOND, J. E., & GLOVER, G. H. (1997). Separate neural bases of two fundamental memory processes in the human medial temporal lobe. *Science*, **276**, 264-266.

GLUCK, M. A., & MYERS, C. (1993). Hippocampal mediation of stimulus representation: A computational theory. *Hippocampus*, **3**, 491-516.

GLUCK, M. A., & MYERS, C. (2001). *Gateway to memory: An introduction to neural network modeling of the hippocampus in learning and memory*. Cambridge, MA: MIT Press.

GLUCK, M. A., OLIVER, L. M., & MYERS, C. E. (1996). Late-training amnesic deficits in probabilistic category learning: A neurocomputational analysis. *Learning & Memory*, **3**, 326-340.

GOLINKOFF, M., & SWEENEY, J. A. (1989). Cognitive impairments in depression. *Journal of Affective Disorders*, **17**, 105-112.

GOLOMB, J., KLUGER, A., DE LEON, M. J., FERRIS, S., CONVIT, A., MITTELMAN, M., ET AL. (1994). Hippocampal formation size in normal human aging: A correlate of delayed secondary memory performance. *Learning & Memory*, **1**, 45-54.

GRADY, C. L., MCINTOSH, A. R., HORWITZ, B., MAISOG, J. M., UNGERLEIDER, L. G., MENTIS, M. J., ET AL. (1995). Age-related reductions in human recognition memory due to impaired encoding. *Science*, **269**, 218-221.

GRUNWALD, T., LEHNERTZ, K., HEINZE, H. J., HELMSTAEDTER, C., & ELGER, C. E. (1998). Verbal novelty detection within the human hippocampus proper. *Proceedings of the National Academy of Sciences*, **95**, 3193-3197.

GURECKIS, T., & LOVE, B. C. (2002). Who says models can only do what you tell them? Unsupervised category learning data, fits, and predictions. In W. D. Gray & C. D. Schunn (Eds.), *Proceedings of the 24th Annual Conference of the Cognitive Science Society* (pp. 399-404). Mahwah, NJ: Erlbaum.

GURECKIS, T., & LOVE, B. C. (2003). Towards a unified account of supervised and unsupervised learning. *Journal of Experimental & Theoretical Artificial Intelligence*, **15**, 1-14.

GURECKIS, T., & LOVE, B. C. (2004). Common mechanisms in infant and adult category learning. *Infancy*, **5**, 173-198.

HAY, J. F., & JACOBY, L. L. (1999). Separating habit and recollection in

young and older adults: Effects of elaborative processing and distinctiveness. *Psychology & Aging*, **14**, 122-134.

Heaton, R. K. (1981). *Wisconsin Card Sorting Test manual*. Odessa, FL: Psychological Assessment Resources.

Hess, T. M. (1982). Visual abstraction processes in young and old adults. *Developmental Psychology*, **18**, 473-484.

Hess, T. M., Pullen, S. M., & McGee, K. A. (1996). Acquisition of prototype-based information about social groups in adulthood. *Psychology & Aging*, **11**, 179-190.

Hess, T. M., & Slaughter, S. J. (1986). Aging effects on prototype abstraction and concept identification. *Journal of Gerontology*, **41**, 214-221.

Hockley, W. E. (1992). Item versus associative information: Further comparisons of forgetting rates. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 1321-1330.

Holdstock, J. S., Mayes, A. R., Cezayirli, E., Isaac, C. L., Aggleton, J. P., & Roberts, N. (2000). A comparison of egocentric and allocentric spatial memory in a patient with selective hippocampal damage. *Neuropsychologia*, **38**, 410-425.

Johnson, M. H. (2003). Development of human brain functions. *Biological Psychiatry*, **54**, 1312-1316.

Kéri, S. (2003). The cognitive neuroscience of category learning. *Brain Research Reviews*, **43**, 85-109.

Kirchhoff, B. A., Wagner, A. D., Maril, A., & Stern, C. E. (2000). Prefrontal-temporal circuitry for episodic encoding and subsequent memory. *Journal of Neuroscience*, **20**, 6137-6180.

Klinger, L. G., & Dawson, G. (2001). Prototype formation in autism. *Development & Psychopathology*, **13**, 111-124.

Knight, R. T. (1996). Contribution of human hippocampal region to novelty detection. *Nature*, **383**, 256-259.

Knowlton, B. J., & Squire, L. R. (1993). The learning of categories: Parallel brain systems for item memory and category knowledge. *Science*, **262**, 1747-1749.

Kretschmann, J. J., Kammradt, G., Krauthausen, I., Sauer, B., & Wingert, F. (1986). Growth of the hippocampal formation in man. *Bibliotheca Anatomica*, **28**, 27-52.

Kroll, N. E. A., Knight, R. T., Metcalfe, J., Wolf, E. S., & Tulving, E. (1996). Cohesion failure as a source of memory illusions. *Journal of Memory & Language*, **35**, 176-196.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, **99**, 22-44.

Landfield, P. W., Baskin, R. K., & Pitler, T. A. (1981). Brain aging correlates: Retardation by hormonal-pharmacological treatments. *Science*, **214**, 581-584.

Li, S., Naveh-Benjamin, M., & Lindenberger, U. (2005). Aging neuromodulation impairs associative binding: A neurocomputational account. *Psychological Science*, **16**, 445-450.

Logan, G. D. (1988). Towards an instance theory of automatization. *Psychological Review*, **95**, 492-527.

Love, B. C. (2002a, November). *Aging effects in category learning*. Paper presented for the Mind, Brain, and Behavior Forum Series, Harvard University, Cambridge, MA.

Love, B. C. (2002b). Comparing supervised and unsupervised category learning. *Psychonomic Bulletin & Review*, **9**, 829-835.

Love, B. C. (2003, September). *Infants, amnesiacs, and the MTL*. Paper presented at the J. S. McDonnell Foundation Cognitive Neuroscience of Category Learning Meeting, New York.

Love, B. C. (2005). Environment and goals jointly direct category acquisition. *Current Directions in Psychological Science*, **14**, 195-199.

Love, B. C., & Gureckis, T. M. (2004). The hippocampus: Where a cognitive model meets cognitive neuroscience. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the Twenty-Sixth Annual Conference of the Cognitive Science Society* (p. 1589). Mahwah, NJ: Erlbaum.

Love, B. C., & Gureckis, T. M. (2005). Modeling learning under the influence of culture. In W. Ahn, R. L. Goldstone, B. C. Love, A. B. Markman, & P. Wolff (Eds.), *Categorization inside and outside the laboratory: Essays in honor of Douglas L. Medin* (pp. 229-248). Washington, DC: American Psychological Association.

Love, B. C., Gureckis, T. M., & Worchel, J. (2007). *Category learning and aging: No exception is the rule*. Manuscript submitted for publication.

Love, B. C., Markman, A. B., & Yamauchi, T. (2000). Modeling classification and inference learning. In *Proceedings of the Seventeenth National Conference on Artificial Intelligence* (pp. 136-141). Menlo Park, CA: AAAI Press.

Love, B. C., & Medin, D. L. (1998a). Modeling item and category learning. In S. J. Derry & M. A. Gernsbacher (Eds.), *Proceedings of the 20th Annual Conference of the Cognitive Science Society* (pp. 639-644). Mahwah, NJ: Erlbaum.

Love, B. C., & Medin, D. L. (1998b). SUSTAIN: A model of human category learning. In *Proceedings of the Fifteenth National Conference on Artificial Intelligence* (pp. 671-676). Menlo Park, CA: AAAI Press.

Love, B. C., Medin, D. L., & Gureckis, T. (2004). SUSTAIN: A network model of human category learning. *Psychological Review*, **111**, 309-332.

Lupien, S. L., de Leon, M., de Santi, S., Convit, A., Tarshish, C., Nair, N. P. V., et al. (1998). Cortisol levels during human aging predict hippocampal atrophy and memory deficits. *Nature Neuroscience*, **1**, 69-73.

Maguire, E. A., Gadian, D. G., Johnsrude, I. S., Good, C. D., Ashburner, J., Frackowiak, R. S. J., & Frith, C. D. (2000). Navigation-related structural change in the hippocampi of taxi drivers. *Proceedings of the National Academy of Sciences*, **97**, 4398-4403.

Mangels, J. A. (1997). Strategic processing and memory for temporal order in patients with frontal lobe lesions. *Neuropsychology*, **11**, 207-221.

Mayes, A. R., Isaac, C. L., Holdstock, J. S., Hunkin, N. M., Montaldi, D., Downes, J. J., et al. (2001). Memory for single items, word pairs, and temporal order in a patient with selective hippocampal lesions. *Cognitive Neuropsychology*, **18**, 97-123.

McClelland, J., McNaughton, B., & O'Reilly, R. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, **102**, 419-457.

McEwen, B. S., Weiss, J. M., & Schwartz, L. S. (1968). Selective retention of corticosterone by limbic structures in rat brain. *Nature*, **220**, 911-912.

McIntyre, J. S., & Craik, F. I. M. (1987). Age differences in memory for item and source information. *Canadian Journal of Psychology*, **41**, 175-192.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, **85**, 207-238.

Merzenich, M. M., & Sameshima, K. (1993). Cortical plasticity and memory. *Current Opinion in Neurobiology*, **3**, 187-196.

Mitchell, K. J., Johnson, M. K., Raye, C. L., & D'Esposito, M. (2000). fMRI evidence of age-related hippocampal dysfunction in feature binding in working memory. *Cognitive Brain Research*, **10**, 197-206.

Montaldi, D., Spencer, T., Roberts, N., & Mayes, A. (2006). The neural system that mediates familiarity memory. *Hippocampus*, **16**, 504-520.

Naveh-Benjamin, M. (2000). Adult age differences in memory performance: Tests of an associative deficit hypothesis. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26**, 1170-1187.

Naveh-Benjamin, M., Hussain, Z., Guez, J., & Bar-On, M. (2003). Adult age differences in episodic memory: Further support for an associative-deficit hypothesis. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **29**, 826-837.

Nelson, C. A. (1995). The ontogeny of human memory: A cognitive neuroscience perspective. *Developmental Psychology*, **31**, 723-738.

Norman, K. A., & O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: A complementary learning systems approach. *Psychological Review*, **110**, 611-646.

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, **115**, 39-57.

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review*, **101**, 53-79.

Nosofsky, R. M., & Zaki, S. F. (1998). Dissociations between categorization and recognition in amnesic and normal individuals. *Psychological Science*, **9**, 247-255.

Palmeri, T. J., & Flanery, M. A. (1999). Learning about categories in the absence of training: Profound amnesia and the relationship between perceptual categorization and recognition memory. *Psychological Science*, **10**, 526-530.

Parker, A., Wilding, E. L., & Akerman, C. (1998). The von Restorff effect in visual object recognition memory in humans and monkeys: The role of frontal/perirhinal interaction. *Journal of Cognitive Neuroscience*, **10**, 691-703.

Poldrack, R. A., Clark, J., Pare-Blagoev, E. J., Shohamy, D., Creso Moyano, J., Myers, C., & Gluck, M. A. (2001). Interactive memory systems in the human brain. *Nature*, **414**, 546-550.

Posner, M. I., & Dehaene, S. (1994). Attentional networks. *Trends in Neurosciences*, **17**, 75-79.

Ranganath, C., & Knight, R. T. (2003). Prefrontal cortex and episodic memory: Integrating findings from neuropsychology and event-related functional neuroimaging. In A. Parker, E. L. Wilding, & T. J. Bussey (Eds.), *The cognitive neuroscience of memory: Encoding and retrieval* (pp. 83-99). Philadelphia: Psychology Press.

Ranganath, C., & Rainer, G. (2003). Neural mechanisms for detecting and remembering novel events. *Nature Reviews Neuroscience*, **4**, 193-202.

Ranganath, C., Yonelinas, A. P., Cohen, M. X., Dy, C. J., Tom, S. M., & D'Esposito, M. (2004). Dissociable correlates of recollection and familiarity within the medial temporal lobes. *Neuropsychologia*, **42**, 2-13.

Raz, N. (2000). Aging of the brain and its impact on cognitive performance: Integration of structural and functional findings. In F. I. M. Craik & T. A. Salthouse (Eds.), *Handbook of aging and cognition* (Vol. 2, pp. 1-90). Mahwah, NJ: Erlbaum.

Reber, P. J., Stark, C. E. L., & Squire, L. R. (1998). Cortical areas supporting category learning identified using functional MRI. *Proceedings of the National Academy of Sciences*, **95**, 747-750.

Reed, J. M., Squire, L. R., Patalano, A. L., Smith, E. E., & Jonides, J. (1999). Learning about categories that are defined by object-like stimuli despite impaired declarative memory. *Behavioral Neuroscience*, **113**, 411-419.

Rosch, E., & Mervis, C. B. (1975). Family resemblences: Studies in the internal structure of categories. *Cognitive Psychology*, **7**, 573-605.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, **323**, 533-536.

Sakamoto, Y., & Love, B. C. (2004). Schematic influences on category learning and recognition memory. *Journal of Experimental Psychology*: *General*, **33**, 534-553.

Sapolsky, R. M., Krey, L. C., & McEwen, B. S. (1986). The neuroendocrinology of stress and aging: The glucocorticoid cascade hypothesis. *Endocrine Reviews*, **7**, 284-301.

Save, E., Poucet, B., Foreman, N., & Buhot, M.-C. (1992). Object exploration and reactions to spatial and nonspatial changes in hooded rats following damage to parietal cortex or hippocampal formation. *Behavioral Neuroscience*, **106**, 447-456.

Schacter, D. L., & Buckner, R. L. (1998). Priming and the brain. *Neuron*, **20**, 185-195.

Schmajuk, N., & DiCarlo, J. (1992). Stimulus configuration, classical conditioning, and hippocampal function. *Psychological Review*, **99**, 268-305.

Sears, L. L., & Steinmetz, J. E. (1990). Acquisition of classically conditioned-related activity in the hippocampus is affected by lesions of the cerebellar interpositus nucleus. *Behavioral Neuroscience*, **104**, 681-692.

Seress, L., & Mrzljak, L. (1992). Postnatal development of mossy cells in the human dentate gyrus: A light microscopic Golgi study. *Hippocampus*, **2**, 127-141.

Shors, T. J., Miesegaes, G., Beylin, A., Zhao, M., Rydel, T., & Gould, E. (2001). Neurogenesis in the adult is involved in the formation of trace memories. *Nature*, **410**, 372-376.

Smith, J. D. (2002). Exemplar theory's predicted typicality gradient can be tested and disconfirmed. *Psychological Science*, **13**, 437-442.

Smith, J. D., & Minda, J. P. (2001). Journey to the center of the category: The dissociation in amnesia between categorization and recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 984-1002.

Spencer, W. D., & Raz, N. (1995). Differential effects of aging on memory for content and context: A meta-analysis. *Psychology & Aging*, **10**, 527-539.

Stark, C. E. L., Bayley, P. J., & Squire, L. R. (2002). Recognition memory for single items and for associations is similarity impaired following damage to the hippocampal region. *Learning & Cognition*, **9**, 238-242.

Stern, C. E., Corkin, S., Gonzalez, R. G., Guimaraes, A. R., Baker, J. R., Jennings, P. J., et al. (1996). The hippocampal formation participates in novel picture encoding: Evidence from functional magnetic resonance imaging. *Proceedings of the National Academy of Sciences*, **93**, 8660-8665.

Sutherland, R. J., McDonald, R. J., Hill, C. R., & Rudy, J. W. (1989). Damage to the hippocampal formation in rats selectively impairs the ability to learn cue relationships. *Behavioral & Neural Biology*, **52**, 331-356.

Titov, N., & Knight, R. G. (1997). Adult age differences in controlled and automatic memory processing. *Psychology & Aging*, **12**, 565-573.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology*, **26**, 1-12.

van Praag, H., Christie, B. R., Sejnowski, T. J., & Gage, F. H. (1999). Running enhances neurogenesis, learning, and long-term potentiation in mice. *Proceedings of the National Academy of Sciences*, **96**, 13427-13431.

van Praag, H., Schinder, A. F., Christie, B. R., Toni, N., Palmer, T. D., & Gage, F. H. (2002). Functional neurogenesis in the adult hippocampus. *Nature*, **415**, 1030-1034.

Vargha-Khadem, F., Gadian, D. G., Watkins, K. E., Connelly, A., Van Paesschen, W., & Mishkin, M. (1997). Differential effects of early hippocampal pathology on episodic and semantic memory. *Science*, **277**, 376-380.

Wan, H., Aggleton, J. P., & Brown, M. W. (1999). Different contributions of the hippocampus and perirhinal cortex to recognition memory. *Journal of Neuroscience*, **19**, 1142-1148.

Wirth, S., Yanike, M., Frank, L. M., Smith, A. C., Brown, E. N., & Suzuki, W. A. (2003). Single neurons in the monkey hippocampus and learning of new associations. *Science*, **300**, 1578-1581.

Wixted, J. T. (2005). A theory about why we forget what we once knew. *Current Directions in Psychological Science*, **14**, 6-9.

Yonelinas, A. P., Kroll, N. E. A., Dobbins, I. [G.], Lazzara, M. [M.], & Knight, R. T. (1998). Recognition and familiarity deficits in amnesia: Convergence of remeber–know, process dissociation, and receiver operating characteristic data. *Neuropsychology*, **12**, 323-339.

Yonelinas, A. P., Kroll, N. E. A., Quamme, J. R., Lazzara, M. M., Sauvé, M.-J., Widaman, K. F., & Knight, R. T. (2002). Effects of extensive temporal lobe damage or mild hypoxia on recollection and familiarity. *Nature Neuroscience*, **5**, 1236-1241.

Young, A. H., Sahakian, B. J., Robbins, T. W., & Cowen, P. J. (1999). The effects of chronic administration of hydrocortisone on cognitive function in normal male volunteers. *Psychopharmacology*, **145**, 260-266.

Younger, B., & Cohen, L. B. (1986). Developmental change in infants' perception of correlations among attributes. *Child Development*, **57**, 803-815.

Zaki, S. R. (2004). Is categorization performance really intact in amnesia? A meta-analysis. *Psychonomic Bulletin & Review*, **11**, 1048-1054.

Zaki, S. R., Nosofsky, R. M., Jessup, N. M., & Unverzagt, F. W. (2003). Categorization and recognition performance of a memory-impaired group: Evidence for single-system models. *Journal of the International Neuropsychological Society*, **9**, 394-406.

Zeineh, M. M., Engel, S. A., Thompson, P. M., & Bookheimer, S. Y. (2003). Dynamics of the hippocampus during encoding and retrieval of face–name pairs. *Science*, **299**, 577-580.

NOTES

1. Forming one cluster is consistent with a lack of hippocampal function, because one cluster alone is not sufficient for encoding conjunctive information. Possessing one cluster allows for the encoding

of feature frequencies, but not feature relations. In terms of mapping a single cluster representation to the brain, likely candidates for visual category-learning tasks are a fluency-based responding mechanism, as evidenced by posterior occipital deactivations (Reber et al., 1998), and the elemental-learning system discussed in Fanselow (1999).

2. Although it is an effective demonstration, Palmeri and Flanery (1999) have noted methodological flaws in Knowlton and Squire's (1993) experiments that raise alternative explanations of their findings. Subsequent work improving upon Knowlton and Squire's methodology bolsters the claim that MTL-impaired populations can learn category structures consisting of an underlying prototype (Bozoki et al., 2006).

3. The variability across simulations is due to the interaction of ordering effects with SUSTAIN's incremental recruitment of clusters. Different simulations involve different random presentation orders, which can lead to different cluster solutions. SUSTAIN's predictions for such ordering effects have been systematically explored by Gureckis and Love (2002) with human participants and confirmed. Because of variability across simulations, the results of numerous simulations are averaged together.

At first glance, it might seem odd for amnesic patients to be able to recruit multiple clusters in the recognition task, much as it seems odd that Knowlton and Squire's (1993) amnesic patients displayed above-chance recognition performance. One possibility is that a subset of Knowlton and Squire's patients had some preserved function, since the group was of mixed etiology. Another possibility, previously discussed, is that the hippocampus may be specialized for forming conjunctive codes but other regions may also perform this function (perhaps with less effectiveness), particularly in cases in which the hippocampus is damaged. Gluck and Myers (2001) reviewed evidence in support of this position in the second chapter of their book.

4. The consistent test item shown in Table 1 was itself a study item, raising possible concerns. Experiments 3 and 4 in Younger and Cohen (1986) showed the same pattern of results when this item was removed from the study set. SUSTAIN successfully fits all the experiments contained in Younger and Cohen (Gureckis & Love, 2004).

5. Although not central to our hypothesis, a third test item was also included that contained novel values on all three attributes. This item received the longest looking times for both groups of infants and was included primarily to ensure that the infants were encoding the attributes of the study items in a manner consistent with the experimenters' expectations. SUSTAIN also predicts that this novel item should elicit the longest looking times for both groups.

6. If other model parameters, in addition to the key parameter, related to memory fidelity are allowed to vary across simulations of different populations, both models are likely to capture the complete data pattern at the cost of explanatory power.

APPENDIX

Here, SUSTAIN's implementation will be described, as well as the procedures used to apply it to the three simulations discussed in the main text.

Input Representation

A nominal stimulus attribute (as used in Younger & Cohen, 1986, and Love et al., 2007), containing k distinct values is represented in the model by k input units. The unit that denotes the value of the attribute is set to one, and all of the other units forming the attribute are set to zero. Thus, a complete stimulus is represented by $I^{\text{pos}_{ik}}$, where i indexes the stimulus attribute and k indexes the nominal values for attribute i. The "pos" in I^{pos} denotes that the current stimulus occupies a particular position in a multidimensional representational space.

The distance μ_{ij} between the ith stimulus attribute and cluster j's position along the ith attribute is defined as

$$\mu_{ij} = \frac{1}{2} \sum_{k=1}^{v_i} \left| I^{\text{pos}_{ik}} - H_j^{\text{pos}_{ik}} \right|, \tag{A1}$$

where v_i is the number of different nominal values on the ith attribute, $I^{\text{pos}_{ik}}$ is the position of the input stimulus on the ith attribute for value k, and $H_j^{\text{pos}_{ik}}$ is cluster j's position on the ith attribute for value k. The distance μ_{ij} is always between 0 and 1, inclusive. Distance calculations for dot pattern stimuli will be discussed in the section detailing the simulation procedures for Knowlton and Squire's (1993) simulations.

Generating a Response

Once the stimulus is encoded, clusters are activated on the basis of their similarity to the input item. The activation of a cluster is given by

$$H_j^{\text{act}} = \frac{\sum_{i=1}^{n_a} \left(\lambda_i\right)^r e^{-\lambda_i \mu_{ij}}}{\sum_{i=1}^{n_a} \left(\lambda_i\right)^r}, \tag{A2}$$

where H_j^{act} is the activation of the jth cluster, n_a is the number of stimulus attributes, λ_i is the tuning of the receptive field (which implements selective attention in SUSTAIN) for the ith input attribute, and r is the attentional parameter (always nonnegative). At the start of learning, λ_i is set to 1.

Clusters compete to respond to input patterns through a process of mutual inhibition. The final output, H_j^{out}, of each cluster j is computed according to

$$H_j^{\text{out}} = \frac{\left(H_j^{\text{act}}\right)^\beta}{\sum_{i=1}^{n_c} \left(H_i^{\text{act}}\right)^\beta} H_j^{\text{act}}, \tag{A3}$$

where n_c is the current number of the clusters and β is a lateral inhibition parameter (always nonnegative) that regulates cluster competition.

APPENDIX (Continued)

In supervised learning, SUSTAIN selects the cluster with the largest output value, and only this cluster is allowed to pass its output, H_j^{out}, across the upper layer of connection weights to the final output units. The winning cluster, H_m, passes its output to the k output units of the unknown (queried) attribute z by

$$C_{zk}^{\text{out}} = w_{m,zk} H_m^{\text{out}}, \tag{A4}$$

where C_{zk}^{out} is the output of the unit representing the kth nominal value of the unknown zth attribute and $w_{m,zk}$ is the weight from the winning cluster, H_m, to output unit C_{zk}. In classification learning, z is the category label. The value C_{zk}^{out} is calculated for each of the k units belonging to queried attribute z.

The probability of making a response k (the kth nominal value) for the queried attribute z is

$$P(k) = \frac{e^{\left(d \cdot C_{zk}^{\text{out}}\right)}}{\sum_{j=1}^{v_z} e^{\left(d \cdot C_{zk}^{\text{out}}\right)}}, \tag{A5}$$

where d is the response parameter (always nonnegative) and v_z is the number of nominal units (and hence, output units) forming the queried attribute z.

Recognition and unsupervised categorization judgments are modeled in an identical fashion. In these cases, SUSTAIN's response is determined by summing the output H_j^{out} for all clusters:

$$R = \sum_{j=1}^{n_c} H_j^{\text{out}}, \tag{A6}$$

where R is the recognition score for the current stimulus. This recognition score is a measure of the model's overall familiarity for a stimulus.

Cluster Recruitment

In all learning tasks, the initial cluster is centered on the first stimulus presentation, and additional clusters are recruited in response to surprising events encountered throughout learning. Newly recruited clusters are initially centered on the surprising stimulus (i.e., all μ_{ij} will be zero for the new cluster and the current stimulus).

In supervised category-learning tasks, feedback is provided after the model makes its response. SUSTAIN attempts to recruit a new cluster when the winning cluster, H_m, predicts the incorrect category label. In unsupervised learning, this signal to create a new cluster is generated whenever the activation of the winning cluster, H_m^{act}, is below the parameter τ_s.

Not all attempts to recruit a new cluster are successful. The ability to construct a new cluster in response to a surprising event is proposed to be hippocampal dependent. Simulations with low hippocampal function will fail to recruit a cluster in response to a surprising stimulus when that stimulus is similar to an existing cluster. Formally, a new cluster is recruited when the activation of the winning cluster, H_m^{act}, is less than the value of the hippocampal function parameter τ_h.

Note that in unsupervised category learning, the lower of the two thresholds, τ_s and τ_h, determines when a cluster can be recruited. Formally, these two parameters could be replaced with a single parameter. Nevertheless, these parameters are conceptually distinct and could be identifiable, given other data sources. For instance, an amnesic patient could be surprised by a stimulus but not be able to recruit a cluster to encode the stimulus. However, because these two parameters are not identifiable in the current simulations, we report the minimum of the two thresholds, $\min(\tau_s,\tau_h)$. On the basis of our literature review and the demands of the learning tasks simulated, we assume that the deficit for the groups considered here lies in hippocampal function (i.e., $\tau_h < \tau_s$), although the same results would hold if the deficit was in the PFC.

Learning

Learning rules are applied on each training trial that update the clusters and weights. On trials on which a new cluster is recruited, it will be selected as the winning (i.e., most activated) cluster, due to the fact that this cluster is centered on the current input item. Otherwise, the cluster most similar to the stimulus will be the winner. For the winning cluster H_m, the position of the cluster is adjusted by

$$\Delta H_m^{\text{pos}_{ik}} = \eta \left(I^{\text{pos}_{ik}} - H_m^{\text{pos}_{ik}} \right), \tag{A7}$$

where η is the learning rate parameter. Thus, the winning cluster moves toward the current stimulus. This learning rule tends to center the cluster amid its members.

Receptive field tunings (which implement attribute attention) are updated according to

$$\Delta\lambda_i = \eta e^{-\lambda_i \mu_{im}} \left(1 - \lambda_i \mu_{im}\right), \tag{A8}$$

where m is the index of the winning cluster. Only the winning cluster updates the value of λ_i.

APPENDIX (Continued)

In supervised category learning, the one-layer delta learning rule (Rumelhart, Hinton, & Williams, 1986) is used to adjust weights from the winning cluster to output units:

$$\Delta w_{m,zk} = \eta\left(t_{zk} - C_{zk}^{\text{out}}\right)H_m^{\text{out}},\tag{A9}$$

where z is the queried attribute (i.e., the category label in classification learning).

Simulation Procedures

In the following section, we will provide simulation details for each reported study. Simulations were conducted in a manner that paralleled the procedures used for human participants. For example, in each study, the number of learning trials and randomization procedures for SUSTAIN simulations matched the experiences of human participants.

Instead of fitting each study individually, we used a single set of core parameters across all three studies. These core parameters ($r = 2.844$, $\beta = 2.386$, $d = 12.000$, and $\eta = .093$) were set to the global parameters reported in Love et al. (2004) that enabled SUSTAIN to correctly predict patterns of human performance across a wide range of studies. Other decisional parameters (which effect quantitative fit, but not qualitative predictions) varied across the three simulations reported here. All of these peripheral parameters will be discussed in the sections below devoted to particular simulations. Of course, the τ_h parameters necessarily varied across studies. Unfortunately, scaling issues with particular stimuli types prevent the direct comparison of τ_h values across studies.

Knowlton and Squire (1993). Stimuli were random dot patterns as shown in Figure 4. Because these stimuli do not contain nominal attributes or features, input to the model was a set of Cartesian coordinates that described the location of each of the nine dots in a 50×50 pixel grid. A complete stimulus was defined by a set of nine tuples, each of which specified the location of one dot in the stimulus pattern. The same representation characterized cluster positions.

The distance between the current stimulus and cluster H_j was determined by first calculating the average distance between pairs of corresponding dots in both patterns. Correspondences between dots were determined by attempting to minimize the average distance over all possible permutations of dot correspondences. Unfortunately, determining these correspondences is computationally demanding, because there are $9! = 362{,}880$ ways to map the nine dots, all of which have to be searched to find the minimum average distance on each trial for each cluster. To approximate this exhaustive computation, a greedy heuristic was used that placed two dots in correspondence one at a time on the basis of the closest pairs. At each step of the greedy algorithm, the two dots with the shortest Euclidean distance between them were placed in correspondence and then removed from consideration. This simple algorithm was able to recover the minimal mapping (out of 362,880 possible mappings) 33% of the time. Furthermore, nonoptimal correspondences closely approximated optimal solutions. For example, we computed the distance between 10 test patterns against 10 random training patterns (100 total comparisons) and found that the true average distance for this set was 4.3, whereas the greedy method yielded an average distance of 4.8. In comparison, random dot correspondences yield an average distance of 11.0.

Following Smith and Minda (2001), the log of this average dot distance plus 1 (to ensure positive distances) served as our psychological distance measure μ_{1j}, replacing Equation A1 which was used in the case of nominal-valued attributes. With only μ_{1j} serving as input, Equation A2 reduces to

$$H_j^{\text{act}} = e^{-\lambda_c \mu_{1j}},\tag{A10}$$

where H_j^{act} is the activation of the jth cluster and λ_c is a dot pattern generalization parameter that controls the gradient of generalization in the model. In the best-fit results shown in Figures 5 and 6, there was no learning on this dot pattern generalization parameter, and it was treated as a parameter that differed between the categorization ($\lambda_c = 0.8$) and recognition ($\lambda_c = 1.3$) tasks. Critically, amnesic patients were modeled as having lower setting of the hippocampal parameter, relative to controls [$\min(\tau_s,\tau_h) = .053$ vs. .13, respectively].

Both categorization and recognition responses in the model were computed using the following response function:

$$P(\text{endorsement}) = \frac{R}{R + k},\tag{A11}$$

where $P(\text{endorsement})$ is the probability that the stimulus is judged as old in the recognition task or as a member of the category in the categorization task. The recognition score, R, is computed according to Equation A6. The response criteria parameter, k, was held constant ($k = .19$) across tasks. Thus, there was one parameter [$\min(\tau_s,\tau_h)$] that differed between the simulated groups and another (λ_c) that differed between tasks; otherwise, each simulated condition was identical. In this sense, our approach to simulating this data set bears a strong resemblance to previous efforts (Nosofsky & Zaki, 1998; Smith & Minda, 2001). However, we also considered the ability of SUSTAIN to account for these data using only a single parameter, $\min(\tau_s,\tau_h)$, by allowing it to learn the value of λ_c over the course of training according to Equation A8. Given an initial setting of $\lambda_c = 0.47$ and $k = .26$ across all tasks and populations allowed the model to provide a similar fit [in these simulations, $\min(\tau_s,\tau_h)$ was .24 and .27 for amnesics and controls, respectively].

APPENDIX (Continued)

Younger and Cohen (1986). Four-month-old infants were modeled as having lower setting of the hippocampal function parameter, relative to 10-month-olds [$\min(\tau_s,\tau_h)$ = .10 vs. .67]. All other parameters were shared by both groups. At the end of the study phase, looking time was predicted by SUSTAIN's recognition score (see Equation A6). More familiar items were assumed to predict shorter looking times. Absolute looking time (in seconds) was calculated by linearly regressing mean infant looking times with the average recognition scores [R^2 = .96; $F(1,2)$ = 47.5, $p < .03$].

Love, Gureckis, and Worchel (2007). Older adults were modeled as having lower setting of the hippocampal function parameter, relative to younger adults (τ_h = .07 vs. 1.00). All other parameters were shared by both groups. The parameter τ_s is not germane, because surprising events were mispredictions in supervised learning. One additional parameter, $\lambda_{distinct}$, was necessary for these simulations, due to the fact that the rule-relevant dimension was cued throughout the study phase and, therefore, was likely to be more salient to participants. Attention on this particular attribute was fixed ($\lambda_{distinct}$ = 5.1) throughout the simulations instead of being initially set to 1 and updated by Equation A8.