INFANCY, *5*(2), 173–198

Common Mechanisms in Infant and Adult Category Learning

Todd M. Gureckis and Bradley C. Love
Department of Psychology
University of Texas at Austin

Computational models of infant categorization often fail to elaborate the transitional mechanisms that allow infants to achieve adult performance. In this article, we apply a successful connectionist model of adult category learning to developmental data. The Supervised and Unsupervised Stratified Adaptive Incremental Network (SUSTAIN) model is able to account for the emergence of infants' sensitivity to correlated attributes (e.g., has wings and can fly). SUSTAIN offers 2 complimentary explanations of the developmental trend. One explanation centers on memory storage limitations, whereas the other focuses on limitations in perceptual systems. Both explanations parallel published findings concerning the cognitive and sensory limitations of infants. SUSTAIN's simulations suggest that conceptual development follows a continuous and smooth trajectory despite qualitative changes in behavior and that the mechanisms that underlie infant and adult categorization might not differ significantly.

The ability to form categories of distinct objects is of critical importance in human cognition. Consequently, there has been a keen interest in understanding how this ability develops. Following a strategy that has proven successful in the adult categorization literature, computational models have been applied to infant learning data to understand the underlying mechanisms (Mareschal & French, 2000; Mareschal, French, & Quinn, 2000; Quinn & Johnson, 1997; Shultz, 2001).

Although general developmental principles have emerged, it is unclear how current models of infant categorization can explain the transition to adult competency. Models built to capture infant performance are often inappropriate for modeling adults. For example, Mareschal and French's (1997) autoencoder network

Requests for reprints should be sent to Todd M. Gureckis, Department of Psychology, 1 University Station #A8000, University of Texas, Austin, TX 78712. E-mail: gureckis@love.psy.utexas.edu

model successfully accounts for catastrophic interference findings in the developmental literature, but adults do not demonstrate this kind of memory interference (Ratcliff, 1990), and it is unclear how one might extend their model to address this fact. A complete developmental account of categorization must detail the path from infant to adult performance (cf. Cohen, Chaput, & Cashon, 2002; Sirois & Shultz, 1998).

In this article, we apply a successful model of adult learning, the Supervised and Unsupervised Stratified Adaptive Incremental Network (SUSTAIN) model (Love, Medin, & Gureckis, 2003), to Younger and Cohen's (1986) infant learning data. Although this is SUSTAIN's first application to developmental data, the model has accounted for an array of challenging adult 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), and unsupervised learning (Gureckis & Love, 2002a, 2002b, 2003).

Despite its newness as a developmental model, SUSTAIN offers some unique advantages for modeling such data. One advantage is that SUSTAIN is situated in the experimental task and procedure. SUSTAIN is a trial-by-trial model of category learning (i.e., one human learning trial equals one model learning trial). SUSTAIN begins with a single cluster (clusters are akin to hidden units) and only adds more in response to a surprising event that happens on a particular learning trial. In unsupervised learning situations, SUSTAIN's output is easily related to infant looking time.

In comparison, other models are more divorced from the experimental context. For example, other models update connection weights in batch mode after a series of learning trials instead of on a trial-by-trial basis (e.g., Mareschal & French, 2000; Sirois & Shultz, 1998). In addition to time shifting error correction, the number of learning trials experienced in Sirois and Shultz's (1998) cascade correlation model depends on which age group is being modeled even for cases in which all age groups in the actual experiment receive an equal number of learning trials.

Many connectionist models begin with a somewhat arbitrarily chosen and fixed architecture (i.e., the number of hidden units is chosen by the modeler and does not change during learning). The motivation for a particular architectural choice is not always clear, nor is how such models might develop to account for adult competencies. However, this is not generally a concern for SUSTAIN or cascade correlation models (Fahlman & Lebiere, 1990), which generate their network architecture during learning.

The way in which hidden units are recruited in SUSTAIN and cascade correlation differs considerably. In cascade correlation, hidden units are placed in a cascading fashion so that each hidden unit receives input from all of the previous hidden and input units (Shultz, Schmidt, Buckingham, & Mareschal, 1995). The primary goal of such architectural changes is to rapidly minimize overall network error. In contrast, SUSTAIN adds new clusters to a single layer of its network in an attempt to optimally represent the stimulus space. Thus, architectural changes in SUSTAIN are easily interpreted as changes in category representation. Clusters can be identified at the end of a simulation as representing common category landmarks such as exceptions, prototypes, and subprototypes.

In the remainder of this article, we introduce SUSTAIN, review Younger and Cohen's (1986) key findings, evaluate SUSTAIN's account of these data, and consider the implications of our results. SUSTAIN offers two complimentary explanations of the developmental progression observed by Younger and Cohen. Both explanations are physiologically inspired and posit that infants have reduced sensitivity to stimulus differences. One explanation focuses on the ability to form conjunctive codes in memory, and the other explanation centers on perceptual limitations. In each case, the developmental changes in SUSTAIN are continuous, even when SUSTAIN's behavior suggests distinct developmental stages.

DESCRIPTION OF SUSTAIN

We begin our introduction to SUSTAIN by presenting an overview of the operation of the model. This is followed by a discussion of the key psychological principles from which the model is derived. Finally, we discuss some advantages SUSTAIN offers for modeling infant learning data.

This introduction serves to highlight the most important features of the model and provides sufficient background to interpret the simulation results. The Appendix details the mathematical equations that follow from SUSTAIN's general principles.

Overview of SUSTAIN

SUSTAIN is a network model of human category learning. Figure 1 shows a graphical overview of the model. On each learning trial, SUSTAIN takes as input a description of the current stimulus item represented to the model by a set of perceptual feature dimensions. For example, the large, purple square at the bottom of Figure 1 is represented to the model by the feature dimensions color, size, and stripe. Like other models of category learning (e.g., Anderson, 1991), SUSTAIN treats the category membership (or category label) of a stimulus item as simply another stimulus feature dimension. In Figure 1, SUSTAIN is being asked to predict which category the current stimulus belongs to, thus there is a "?" over the category label dimension. Instead of being asked to predict the category label, SUSTAIN could also be asked to predict the color of a stimulus given its size, stripe, and category membership. This flexible strategy allows SUSTAIN to model inference tasks.

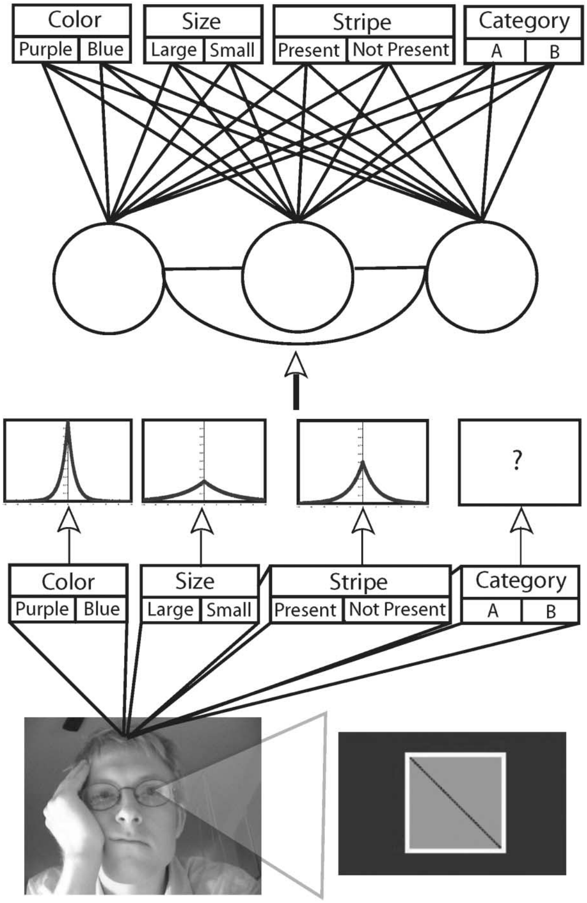

FIGURE 1 An overview of the SUSTAIN model.

SUSTAIN maintains a selective attention mechanism that allows it to learn to focus attention on stimulus dimensions that are particularly useful for the current categorization task (similar to Krucshcke, 1992). In Figure 1, this attentional mechanism is illustrated by the exponentially shaped receptive fields positioned above each input dimension.

The internal representations in the model consist of a set of clusters (denoted by the circles in the center of Figure 1). Categories are represented in the model as one or more associated clusters. Initially, the network has only one cluster that is centered on the first input pattern. As new stimulus items are presented, the model attempts to assign these new items to an existing cluster. This assignment is done through an unsupervised procedure based on the similarity of the new item to the stored clusters.

When a new item is assigned to a cluster, the cluster updates its internal representation to become the average of all items assigned to the cluster so far.

However, if SUSTAIN discovers through feedback that this similarity-based assignment is incorrect, a new cluster is created to encode the current item as an exception (for a concrete example of this, see Principle 3 in the following section). In unsupervised learning tasks there is no corrective feedback, so instead SUSTAIN creates a new cluster if the current stimulus item is not sufficiently similar to any existing clusters (the threshold for this sufficiency is controlled by a parameter in the model). Both of these cluster recruitment strategies are unified under the principle of "adaptation to surprise" (Gureckis & Love, 2003). In supervised learning, SUSTAIN creates a new cluster in response to a surprising misclassification, whereas in unsupervised learning, a new cluster is created when the model encounters a surprisingly novel stimulus item. Infant category learning studies are typically unsupervised.

Clusters compete with each other to respond to the current stimulus item. The cluster that wins this competition passes its activation over connection weights to a set of output units. These output units replicate the structure of the input dimensions (illustrated at the top of Figure 1). The connection weights are adjusted over the course of learning so that the association between each cluster and the appropriate response for members of that cluster is strengthened. For example, a cluster with members that are mostly in Category A would develop over the course of learning a stronger connection to the Category A output unit than to the Category B output unit. The activation of an output unit is proportional to the strength of the activation passed from the winning cluster and the strength of the connection weight. SUSTAIN's ultimate response is biased toward the most activated output unit. In this way, classification decisions are ultimately based on the cluster to which an instance is assigned.

In many unsupervised learning experiments (including the studies discussed here), there are not contrasting categories into which stimuli are being classified. In these cases, instead of having Category A and Category B, there is only a single global category representing items seen during learning. Consequently, SUSTAIN has only a single output unit representing this global category. The degree to which this unit is activated indicates the level of familiarity the model has for the item. Thus, we refer to this unit as the *category familiarity unit.* The activation of the category familiarity unit should be negatively correlated with infant looking time.

The Key Principles of SUSTAIN

With this general understanding of the operation of the model in mind, we now examine the six key principles of SUSTAIN. These principles highlight the important features of the model and provide the foundation for the model's formalism.

Principle 1: SUSTAIN is directed toward simple solutions. At the start of learning, SUSTAIN has only one cluster that is centered on the first input item. It then adds clusters (i.e., complexity) only as needed to accurately describe the category structure of the learning task. Its selective attention mechanism further serves to bias SUSTAIN toward simple solutions by focusing the model on the stimulus dimensions that provide consistent information.

Principle 2: Similar stimulus items tend to cluster together. In learning to classify stimuli, SUSTAIN will cluster similar items together. For example, different instances of a bird subtype (e.g., sparrows) could cluster together and form a sparrow cluster instead of leaving separate traces in memory for each instance. Clustering is an unsupervised process because cluster assignment is done on the basis of similarity, not feedback.

Principle 3: SUSTAIN relies on both unsupervised and supervised learning processes. As discussed previously, SUSTAIN can cluster based on similarity (an unsupervised process). SUSTAIN's operation is also affected by supervision when available. Consider the example of SUSTAIN learning to classify stimuli as members of the category mammals or birds. Let us assume that a cluster representing four-legged, hairy land creatures has already been acquired by the model, as well as another cluster representing small, winged creatures that fly. The first time SUSTAIN is asked to classify a bat, the model will predict that a bat is a bird because the bat stimulus will be more similar to the existing bird cluster than to the existing mammal cluster. After receiving corrective feedback (supervision), SUSTAIN will note its error and create a new cluster to store the anomalous bat stimulus as an exception. Now, when this bat or one similar to it is presented to SUSTAIN, it will correctly predict that the bat is a mammal. This example also illustrates how SUSTAIN can entertain more complex solutions when necessary through cluster recruitment (see Principle 1).

Principle 4: Clusters are recruited in response to surprising events. As the previous example illustrates, surprising events lead to new clusters being recruited. In unsupervised learning, a surprising event is simply exposure to a stimulus that is not suffciently similar to any existing cluster (i.e., a very novel stimulus).

Principle 5: The pattern of feedback matters. As the bird–mammal example illustrates, feedback affects the inferred category structure. Prediction failures result in a cluster being recruited; thus, different patterns of feedback can lead to different representations being acquired. This principle allows SUSTAIN to predict different acquisition patterns for different learning modes (e.g., inference vs. classification learning) that are informationally equivalent but differ in their pattern of feedback. Likewise, item presentation order in unsupervised learning can affect how items cluster together.

Principle 6: Clusters compete. Clusters can be seen as competing explanations of the input. The strength of the response from the winning cluster (the cluster the current stimulus is most similar to) is attenuated in the presence of other clusters that are somewhat similar to the current stimulus (cf. Sloman's, 1997, account of competing explanations in reasoning).

INFANTS' PERCEPTION OF CORRELATIONS BETWEEN ATTRIBUTES: YOUNGER AND COHEN (1986)

Rosch (1978) argued that natural categories are formed around clusters of correlated attributes. Younger and Cohen (1986) explored infants' ability to acquire categories organized around correlations at 4, 7, and 10 months of age using a habituation technique. The stimuli in their four experiments were pictures of imaginary animals that consisted of three attributes: type of body, type of tail, and type of feet. Each attribute could assume three different values. For example, the type of body of the animal could be similar to either an elephant, a giraffe, or a cow. Younger and Cohen's basic finding was that 4-month-old infants were not sensitive to the correlational structure of the stimulus set, whereas 10-month-old infants were. The performance of 7-month-old infants is not as straightforward to describe but is discussed here along with the results of all four experiments. Infant looking times are shown in Figure 2.

Experiment 1

Experiment 1 assessed the sensitivity of 4- and 7-month-old infants to correlations between stimulus attributes in the absence of independent variation in any other attributes. There were two habituation stimuli in this experiment that had the following abstract structure: 1 1 1 and 2 2 2, where a 1 or 2 represents the particular value of an attribute. For example, the stimulus 1 1 1 might be an item with a elephant body, a horse tail, and club feet, whereas 2 2 2 might be an item with a cow body, a fluffy tail, and hoofed feet. In the habituation phase of the experiment, infants were shown these two stimuli randomly for five blocks (a total of 10 habituation trials).

Three test stimuli (a correlated, an uncorrelated, and a novel item) were created to assess the degree to which infants understood the relation between attributes. The correlated, uncorrelated, and novel test stimuli had the following abstract structure: 2 2 2, 2 1 1, and 3 3 3, respectively. The correlated test item preserved the relation between attributes that was present in the habituation stimuli, whereas the uncorrelated item broke this relation by possessing a combination of feature values

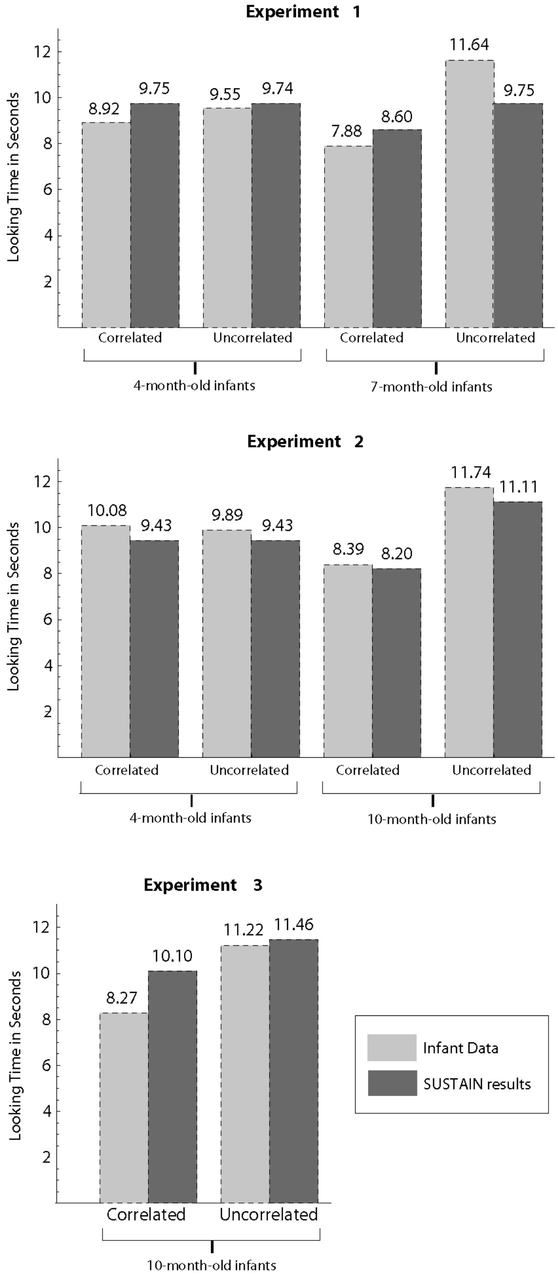

FIGURE 2 A comparison of the infant looking times and SUSTAIN's predicted looking times.

that was not observed during the habituation phase of the experiment. The novel test item had completely new attribute values that the infants had not seen during the habituation phase (this item served as a control). If infants were sensitive to the correlation between features, Younger and Cohen (1986) hypothesized that infants would dishabituate to (i.e., look longer at) the uncorrelated test item than to the correlated test item. However, if the infants were remembering information only about specific features and not relations between features then they would dishabituate only to the novel test item.

In this experiment, the 4-month-old infants looked longer on average at the novel test stimulus than at the correlated and uncorrelated test stimuli. In contrast, the 7-month-old infants looked longer at both the novel and uncorrelated test stimuli relative to the correlated stimulus. The result suggests that 4-month-old infants cannot detect changes in the correlational structure of the test stimuli, whereas 7-month-old infants can.

Experiment 2

In Experiment 2, children were tested at 4, 7, and 10 months of age. Four stimulus items were used in the habituation phase of the experiment, each of which had a perfect correlation between only two of the three possible attributes (1 1 1, 1 1 2, 2 2 1, and 2 2 2). The perfect pairwise correlation between the first and second attribute is here contrasted with independent variation on the third attribute. Infants were shown these four stimuli in a random order for three blocks (for a total of 12 habituation trials). The test items for Experiment 2 were identical to those used in Experiment 1. The correlated and uncorrelated stimuli were again composed of feature values that were seen during the habituation phase of the experiment. The correlated stimulus item preserved the correlational pattern between the first two attributes seen in the habituation phase of the experiment, whereas the uncorrelated stimulus item broke this relation.

As in Experiment 1, the 4-month-old infants dishabituated only to the novel test stimuli. The 10-month-old infants dishabituated to both the novel and uncorrelated test stimuli but not to the correlated test stimulus. In this experiment, the 7-month-old infants never reliably changed their looking behavior to any of the habituation stimuli or the test stimuli.

Experiments 3 and 4

Experiment 3 made a slight modification to the habituation items used in Experiment 2. In Experiment 2, the correlated test item, 2 2 2, was the most similar overall to the habituation items according to the multiplicative similarity rule used in Medin and Schaffer's (1978) context model. Furthermore, the correlated test item was also a habituation item, whereas the uncorrelated test item was novel. Experiment 3 removed the correlated test item from the habituation set, and infants were habituated to the remaining three items. The test item set was the same as Experiment 2.

Infants were tested in only two age groups (7 and 10 months of age). All other aspects of the experiment were identical to Experiment 2. The 7-month-old infants once again failed to habituate in Experiment 3, whereas the 10-month-old infants demonstrated sensitivity to the correlation. In Experiment 4, Younger and Cohen (1986) found that, when trained to a criterion metric, the 7-month-old infants responded much like 4-month-old infants in Experiments 1 and 2.

Summary

The results from all four experiments indicate that 4-month-old children consistently responded to changes in the feature values (the novel stimulus item) but not to changes in the relation between attributes (the uncorrelated stimulus). This finding suggests that 4-month-old infants are unable to integrate and abstract correlations between different attributes. On the other hand, it seems that the ability to process correlations between attributes develops by 10 months of age.

MODELING YOUNGER AND COHEN (1986) WITH SUSTAIN

In all the simulations reported here, the modeling procedure followed the original experimental procedure as closely as possible. Input to the model consisted of a set of attributes or feature dimensions that corresponded to the experimentally manipulated stimulus attributes used by Younger and Cohen (1986; i.e., type of body, type of tail, etc.). Refer to the Appendix for a detailed description of the input coding method. In each experiment, SUSTAIN was given the exact same number of learning or habituation trials (organized into randomized blocks) as the infants. Like infants, SUSTAIN was trained in an unsupervised fashion (i.e., no feedback was provided) during the habituation phase of each experiment. Throughout this phase, the model created new clusters and adjusted its weights on a trial-by-trial basis.

At the end of the habituation phase of each experiment, SUSTAIN was tested for its familiarity toward the correlated and uncorrelated test stimuli. The habituation paradigm assumes that the more familiar something is, the less time an infant will spend looking at it. Thus, looking time is inversely related to the activity of SUSTAIN's category familiarity unit. Absolute looking time was predicted by linearly regressing infant looking time with the activation of this unit.

Simulation 1: The Developmental Limitations in Memory Function Hypothesis

The first set of simulations explores the role of memory limitations on the development of infant categorization ability. Schneider and Bjorklund (1998) argued that infants have a limited memory capacity relative to adults. Such limitations might actually be advantageous to infants over the course of development by working to decrease cognitive load and enabling the bootstrapping of knowledge (Elman, 1993; Turkewitz & Kenny, 1982, 1985).

Memory in SUSTAIN is determined by cluster recruitment. In unsupervised learning, SUSTAIN recruits a new cluster when the current stimulus item is not sufficiently similar (determined by a model parameter, τ) to any existing cluster. The higher the setting of the τ parameter, the more clusters SUSTAIN tends to recruit. Thus, the particular setting of the τ parameter has direct relation to the memory capacity of the model. When the value of τ is low, SUSTAIN collapses all stimulus items into a single global (or prototype) cluster and therefore is only sensitive to feature frequency (like 4-month-old infants). At increasing values of τ, SUSTAIN can recruit multiple clusters that capture the covert category structure in the stimulus set. Thus, SUSTAIN becomes sensitive to correlations between attributes (like 10-month-old infants).

A related hypothesis is that the ability to detect correlations between features requires a representational strategy that can express conjunctions of stimulus features. Prominent theories of hippocampal function assert that one purpose of the hippocampus is to facilitate the creation of "conjunctive codes" that bind together items in episodic memory (Alvarez & Squire, 1994; Marr, 1971; O'Reilly & McClelland, 1994). The traditional view of hippocampal development holds that the hippocampus becomes fully maturated relatively late in infancy (for an opposing position, see Diamond, 1990). Thus, the developmental trends in Younger and Cohen (1986) can be explained by the maturation of the hippocampus, which SUSTAIN models by increasing the value of τ.

Simulations and results. We applied SUSTAIN to the Younger and Cohen (1986) experiments using three different values of τ (one for each age group). The data from the 7-month-old infants in Experiment 2, 3, and 4 were omitted because these infants failed to reliably habituate without a change in experimental procedure. However, the 7-month-old data from Experiment 1 were included (in this experiment the 7-month-old infants did reliably habituate).

The 4-, 7-, and 10-month-old infants had a τ of .10, .54, and .66, respectively. The best-fit value of τ was found using a nonlinear optimization algorithm that attempted to minimize the correlation between the activation of SUSTAIN's category familiarity unit and infant looking times. As predicted, the value of τ rose with age. All other parameters to the model (see Table 1) were fixed at the values

TABLE 1
SUSTAIN's Parameters for the Studies Considered Here

Function/Adjusts	*Symbol*	*Value*
Learning rate	η	.0966
Cluster competition	β	6.40
Decision consistency	d	1.98
Attentional focus	r	10.0
Threshold	τ	.10, .54, .66
Input noise	*noise*	.55, .53, .10

Note. All parameter values were used in previous studies, except for threshold and input noise, which are free parameters. The values of the threshold and input noise parameters are listed in order for 4-, 7-, and 10-month-old infants.

used in adult studies (Love et al., 2003). The model was run 10,000 times in each condition to ensure statistical significance. Infant looking time was regressed onto the mean activation of SUSTAIN's category familiarity unit for the data points considered, $R^2 = .487$, $F(1, 8) = 7.59$, $p < .05$, allowing the model to predict looking time in each experiment. The results are shown in Figure 2.

SUSTAIN captures the key qualitative results of the study. In addition, SUSTAIN does an admirable job at predicting the actual looking times of the infants. SUSTAIN's familiarity with the novel stimulus item was not reported here, but in all simulations SUSTAIN predicts that the novel item will be looked at the longest by infants. This slight overprediction of looking time could be attributable to a nonlinear relation between familiarity and looking time that cannot be accounted for by the linear regression (i.e., at some point the infant will stop looking at an item no matter how novel it is).

SUSTAIN's explanation of the results. In all cases in which Younger and Cohen (1986) reported that infants responded on the basis of attribute frequency, as opposed attribute correlation, SUSTAIN recruited a single cluster. In contrast, SUSTAIN recruited two clusters in all simulations in which infants demonstrated a sensitivity to the correlational structure of the habituation stimuli. SUSTAIN can only show sensitivity to correlation by recruiting multiple clusters.

The clusters and stimuli in SUSTAIN are located in a multidimensional space (see the Appendix). To gain a better understanding of SUSTAIN's explanation of the data, the stimulus and cluster positions were plotted. Figure 3 represents the spatial arrangement of the habituation stimuli, the test stimuli, and SUSTAIN's clusters in the three-dimensional space defined by three stimulus attributes used in Younger and Cohen (1986). In Experiment 1 there were only two habituation items (1 1 1 and 2 2 2, which are positioned at opposing corners of the cubes in Figure 3). In the 4-month-old condition (shown in the left plot of Figure 3), SUSTAIN created only one cluster to represent both of these habituation stimuli. As described

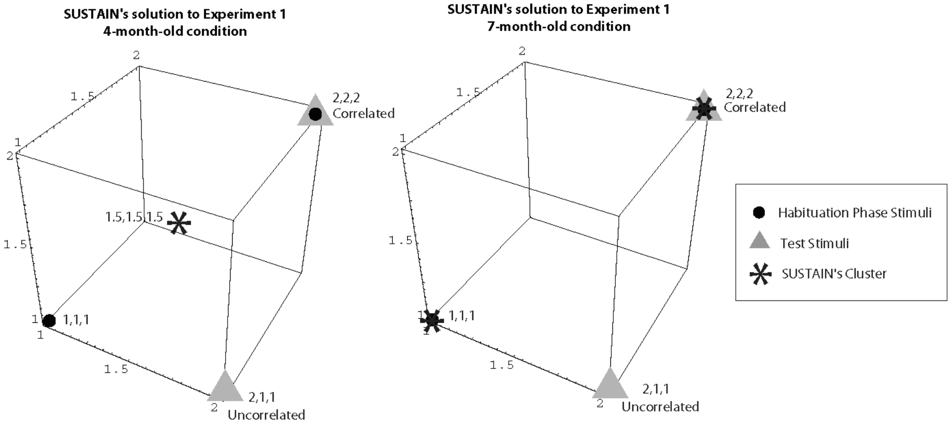

FIGURE 3 A spatial representation of the habituation stimuli, test stimuli, and SUSTAIN's clusters in Experiment 1. Stimuli are represented in this space as three-dimensional points. For example, the habituation stimulus with abstract structure 1 1 1 is represented here as the cartesian point (1,1,1).

earlier, when multiple items are assigned to a single cluster, this cluster moves to become the average of all the stimuli that have been assigned to it. This is shown in the left plot of Figure 3 by SUSTAIN's single cluster (denoted by the star) located at the midpoint between the two habituation stimuli (the point 1.5 1.5 1.5). The correlated and uncorrelated test items are illustrated by gray triangles. From the left plot of Figure 3 it is easy to see that both the correlated and uncorrelated test items are equidistant from SUSTAIN's single cluster. Thus, SUSTAIN predicts equal familiarity for both of these test items (i.e., SUSTAIN will judge both of these items to be equally similar to its internal representation).

In contrast, the right plot of Figure 3 shows SUSTAIN's solution for the 7-month-old infants in Experiment 1. In this set of simulations, SUSTAIN recruited two clusters, one centered on each of the two habituation stimuli (SUSTAIN's clusters are positioned on top of the two habituation stimuli in this illustration). The correlated stimulus item in Experiment 1 (point 2 2 2) has zero distance from one of SUSTAIN's clusters, whereas the uncorrelated stimulus (point 2 1 1) is 1.0 distance units from the nearest cluster. Thus, SUSTAIN predicts that the correlated item will be more familiar than the uncorrelated item.

SUSTAIN's explanation of Experiment 2 unfolds along similar lines. Figure 4 shows the spatial configuration of both SUSTAIN's clusters and the training and test items from Experiment 2. In the 4-month-old condition (shown in the left plot of Figure 4), SUSTAIN recruited a single cluster. This time this single cluster represents the average of the four training items 1 1 1, 1 1 2, 2 2 1, and 2 2 2. However, once again this single cluster is located in the center of the space and is equidistant from both the correlated and uncorrelated test items. Given this configuration,

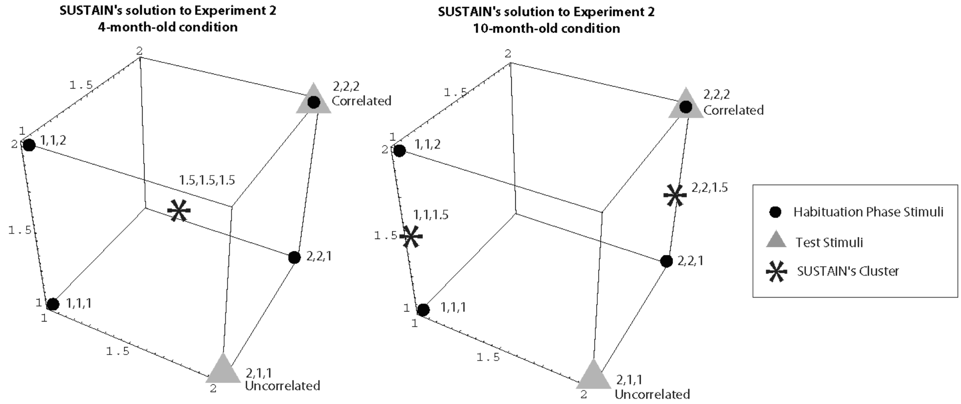

FIGURE 4 A spatial representation of the habituation stimuli, test stimuli, and SUSTAIN's clusters in Experiment 2. Stimuli are represented in this space as three-dimensional points. For example, the habituation stimulus with abstract structure 1 1 1 is represented here as the cartesian point (1,1,1).

SUSTAIN predicts that both the correlated and uncorrelated test items are equally familiar.

In the 10-month-old condition of Experiment 2 (shown in the right plot of Figure 4), SUSTAIN created two clusters. Each of these two clusters represented the average of two of the four habituation 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 at 2 2 1.5). In this case, this uncorrelated test stimulus is farther from the nearest cluster than the correlated test stimulus is. This effect is magnified by SUSTAIN's shift of attention to the two correlation-relevant attributes. A similar solution describes the results of Experiment 3.

Simulation 2: The Degraded Input Encoding Hypothesis

An alternative explanation of the Younger and Cohen (1986) data is that the ability to encode stimulus attributes improves with age (Haynes, White, & Held, 1965). For example, the visual cortex of developing infants is unable to readily perceive high spatial frequencies (Dobson & Teller, 1978; Salaptek & Banks, 1978). The effect of such low-pass filtering of the spatial frequencies is to blur the perception of the retinal image. Like the memory limitations discussed in the previous section, there might be some advantages to perceptual limitations. Recent work suggests that the reduced visual acuity of infants is beneficial for acquiring basic-level concepts (French, Mermillod, Quinn, Chauvin, & Mareschal, 2002).

In addition to the issue of acuity, other developmental differences may affect the ability to encode stimuli. Kelmer (1981) warned that the experimenter-controlled aspects of artificial stimuli might not be represented by infants exactly as the experimenter intends.

In our second set of simulations, we explored the role that coding deficiencies play in infants' categorization ability. Rather than assuming that attribute values are clearly represented, we introduced uncertainty or noise into SUSTAIN's stimulus encoding. The degree of this uncertainty is reduced with increasing age. In the previous simulations, the value of an attribute was clear and certain. For example, an attribute displaying the first value would be represented as [1 0 0]. With input noise set to .5, this same attribute would be represented as [.5 .25 .25]. As the amount of input noise increases, stimuli become more similar to one another.

Simulations and results. In these simulations, we applied SUSTAIN in the same manner as we did in the first set of simulations. The value of the τ parameter was fixed at the value used in the previous simulation for the 10-month-old infants. Developmental differences between the different age groups were modeled by a parameter that controlled the amount of noise in the input encoding. The best-fit value of the noise parameter was once again found using a nonlinear optimization algorithm that attempted to minimize the correlation between the activation of SUSTAIN's category familiarity unit and infant looking time.

The results of the these simulations paralleled the results from the first set of simulations. For brevity, we do not fully present the results. Once again, in all cases in which Younger and Cohen (1986) reported that infants responded on the basis of attribute frequency, as opposed to attribute correlation, SUSTAIN recruited a single cluster. In contrast, SUSTAIN recruited two clusters in all simulations in which infants demonstrated a sensitivity to the correlational structure of the habituation stimuli.

In fact, the two memory and encoding manipulations considered in this article actually predict the same pattern of infant performance. In the first set of simulations, a lower setting of the τ parameter for younger infants accounted for the developmental trends. Lowering the value of the τ parameter decreases the sensitivity of the model to differences between stimuli. In the same way, introducing noise to the input representation reduced a priori differences between stimuli. Thus, at the same setting of the τ parameter, the model is less likely to create additional clusters as the amount of input noise increases. Of course, it would also be possible to account for these data by a combination of either of these two approaches.

GENERAL DISCUSSION

In this article, we have taken a step toward modeling the development of categorization ability from infancy to adulthood. We applied a successful model of adult category learning to infant learning data, and the results were informative.

SUSTAIN captured the developmental trends in Younger and Cohen's (1986) study through gradual changes in its internal parameters. Two compatible development explanations of the data emerged. One explanation focused on the role of memory limitations in the storage and formation of conjunctive codes, whereas the other explanation focused on limitations in stimulus encoding. A unique benefit of this account is that SUSTAIN makes a clear distinction between processes that affect learning in a particular task and those that describe developmental changes between age groups (Thomas & Karmiloff-Smith, in press). However, future research will be needed to tease apart these two candidate explanations of the data.

SUSTAIN's account of these data suggests a number of novel predictions. Although we cannot account for the failure of 7-month-old infants to habituate in the original Younger and Cohen (1986) study, SUSTAIN predicts that the 7-month-old infants were on the cusp of being able to grasp the correlational structure of the habituation items. Manipulations that make this structure more apparent should elevate 7-month-old infants' performance to that of 10-month-old infants in Experiments 2 and 3. One such manipulation is to increase the saliency of attribute values. SUSTAIN forms multiple clusters (which are necessary to capture correlations between attributes) when differences between stimuli are sufficiently large. Both of SUSTAIN's developmental explanations suggest that the threshold for sufficient differences varies over the course of development. Therefore, SUSTAIN predicts that experimentally increasing the differences between attribute values will sensitize 7-month-old infants to correlational structures in the stimulus set.

Another method that should boost 7-month-old infants' performance is to block the presentation order of the habituation stimuli by feature value pairs. In Experiment 1, which consisted of only two stimuli, SUSTAIN correctly predicted that 7-month-old infants would appreciate the correlated structure of the stimulus set. Blocking the presentation order of stimuli in studies like Experiment 2 (which consists of four habituation stimuli) allows SUSTAIN to establish a cluster that captures one attribute value pairing prior to exposure to the second pairing. Once one cluster is firmly established, a stimulus conforming to the second attribute value pairing appears sufficiently different to SUSTAIN to warrant creation of a second cluster, thus capturing the correlational structure of the habituation items. SUSTAIN's trial-by-trial operation allows for such a prediction to be made.

An additional prediction of SUSTAIN is confirmed by Younger (1985). In cases in which SUSTAIN creates one global cluster, the model predicts that the average or prototypical item will be the most familiar item to infants. However, in cases in which SUSTAIN recruits two clusters, SUSTAIN predicts that infants will find the average item much more interesting as it is not very close to either cluster. Younger explored a similar hypothesis with 10-month-old infants. She found that, when presented with an unstructured or broad category, infants had a larger familiarity to the average stimulus than to the modal stimulus item. However, when infants were exposed to a category with two distinct subtypes, infants found the average stimuli to be less familiar.

Despite these successes, one might still ask whether it makes sense to posit a common mechanism for infant and adult category learning. This proposition might not be as unlikely as it seems. There is considerable evidence to suggest that categorization behavior in infants and adults is, in many ways, quite similar. For example, in many cases infants and adults agree on the basic level (Horton & Markman, 1980; Mervis & Crisafi, 1982) and extract the same category prototypes that adults do (Bomba & Siqueland, 1983; Mervis & Crisafi, 1980). Furthermore, Baldwin, Markman, and Melartin (1993) found that 9- to 10-month-old infants make the same kind of inferences from category knowledge that adults do and are capable of inferring nonobvious properties of category members.

One exciting possibility is that infants and adults have the same basic categorization "hardware" and primarily differ in their knowledge or level of domain expertise. This position has been argued for in the analogy literature (Gentner, 1988; Kotovsky & Gentner, 1996). The domain of artificial category learning is ideally suited for exploring this possibility as it typically involves stimuli that lack prior associations. Developmental differences in such tasks may be attributable to parametric differences in memory and perceptual systems and not to qualitative shifts in processing or stagelike progressions.

ACKNOWLEDGMENTS

This work was supported by AFOSR Grant F49620–01–1–0295 to Bradley C. Love. We thank Les Cohen and Andrea Aurthur for helpful comments and discussion during the development of this article.

REFERENCES

Alvarez, P., & Squire, L. (1994). Memory consolidation and the medial temporal lobe: A simple network model. *Proceedings of the National Academy of Science, 91,* 7041–7045.

Anderson, J. (1991). The adaptive nature of human categorization. *Psychological Review, 98,* 409–429.

Baldwin, D., Markman, E., & Melartin, R. (1993). Developmental differences in the acquisition of basic and superordinate categories. *Child Development, 64,* 711–728.

Bomba, P., & Siqueland, E. (1983). The nature and structure of infant form categories. *Journal of Experimental Child Psychology, 35,* 294–328.

Carpenter, G. A., & Grossberg, S. (1987). A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics, and Image Proceedings, 37,* 54–115.

Clapper, J. P., & Bower, G. H. (1991). Learning and applying category knowledge in unsupervised domains. *The Psychology of Learning and Motivation, 27,* 65–108.

Cohen, L. B., Chaput, H. H., & Cashon, C. H. (1985). The segregation of items into categories by ten-month-old infants. *Cognitive Development, 56,* 1574–1583.

Cohen, L. B., Chaput, H. H., & Cashon, C. H. (2002). A constructivist model of infant cognition. *Cognitive Development, 100,* 1–21.

Diamond, A. (1990). Rate of maturation of the hippocampus and the developmental progression of children's performance on the delayed non-matching to sample and visual paired comparison tasks. *Annals of the New York Academy of Sciences, 608,* 394–426.

Dobson, V., & Teller, D. Y. (1978). Visual acuity in human infants: A review and comparison of behavioral and electrophysiological studies. *Vision Research, 18,* 1469–1483.

Elman, J. (1993). Learning and development in neural networks: The importance of starting small. *Cognition, 48,* 71–99.

Fahlman, S. E., & Lebiere, C. (1990). The cascade-correlation learning architecture. In D. S. Touretzky (Ed.), *Advances in neural information processing systems 2: Proceedings of the 1989 conference* (pp. 524–532). San Mateo, CA: Morgan Kaufmann.

French, R., Mermillod, M., Quinn, P., Chauvin, A., & Mareschal, D. (2002). The importance of starting blurry: Simulating improved basic-level category learning in infants due to weak visual acuity. In *Proceedings of the 24th Annual Conference of the Cognitive Science Society* (pp. 322–327). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Gentner, D. (1988). Metaphor as structure mapping: The relational shift. *Child Development, 59,* 47–59.

Gureckis, T., & Love, B. C. (2002a). Modeling unsupervised learning with SUSTAIN. In *Proceedings of the 15th Annual Flairs Conference* (pp. 163–167). CITYXX, FL: XXPUBLISHER NAME.

Gureckis, T., & Love, B. C. (2002b). Who says models can only do what you tell them? Unsupervised category learning data, fits, and predictions. In *Proceedings of the 24th Annual Conference of the Cognitive Science Society* (pp. 399–404). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

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

Hartigan, J. A. (1975). *Clustering algorithms.* New York: Wiley.

Haynes, H., White, B., & Held, R. (1965). Visual accommodation in human infants. *Science, 148,* 528–530.

Horton, M., & Markman, E. (1980). Developmental differences in the acquisition of basic and superordinate categories. *Child Development, 51,* 708–719.

Kelmer, D. (1981). New issues in the study of infant categorization: A reply to Husaim and Cohen. *Merrill-Palmer Quarterly, 27,* 457–463.

Kohonen, T. (1989). *Self-organization and associative memory* (3rd ed.). Berlin: Springer.

Kotovsky, L., & Gentner, D. (1996). Comparison and categorization in the development of relational similarity. *Child Development, 67,* 2797–2822.

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

Love, B. C., Markman, A. B., & Yamauchi, T. (2000). Modeling classification and inference learning. In *Proceedings of the 15th National Conference on Artificial Intelligence* (pp. 136–141). XXCITY, ST: XXPUBLISHER NAME.

Love, B. C., & Medin, D. L. (1998a). Modeling item and category learning. In *Proceedings of the 20th Annual Conference of the Cognitive Science Society* (pp. 639–644). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Love, B. C., & Medin, D. L. (1998b). SUSTAIN: A model of human category learning. In *Proceedings of the 15th National Conference on Artificial Intelligence* (pp. 671–676). Cambridge, MA: MIT Press.

Love, B. C., Medin, D. L., & Gureckis, T. (2003). *SUSTAIN: A network model of human category learning.* Unpublished manuscript.

Luce, R. D. (1959). *Individual choice behavior: A theoretical analysis.* Westport, CT: Greenwood.

Mareschal, D., & French, R. (1997). A connectionist account of interference effects in early infant memory and categorization. In *Proceedings of the 19th Annual Conference of the Cognitive Science Society* (pp. 484–489). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Mareschal, D., & French, R. (2000). Mechanisms of categorization in infancy. *Infancy, 1,* 59–76.

Mareschal, D., French, R., & Quinn, P. (2000). A connectionist account of asymmetric category learning in early infancy. *Developmental Psychology, 36,* 635–645.

Marr, D. (1971). Simple memory: A theory for archicortex. *Philosophical Transactions of the Royal Society of London Series B, 262,* 23–81.

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

Mervis, C., & Crisafi, M. (1980). Acquisition of basic object categories. *Cognitive Psychology, 12,* 496–522.

Mervis, C., & Crisafi, M. (1982). Order of acquisition of subordinate, basic, and superordinate level categories. *Child Development, 53,* 258–266.

O'Reilly, R., & McClelland, J. (1994). Hippocampal conjunctive encoding, storage, and recall: Avoiding a tradeoff. *Hippocampus, 6,* 661–682.

Quinn, P., & Johnson, M. (1997). The emergence of perceptual category representations in young infants. *Journal of Experimental Child Psychology, 66,* 236–263.

Ratcliff, R. (1990). Connectionist models of recognition memory: Constraints imposed by learning and forgetting functions. *Psychological Review, 97,* 285–308.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

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

Salaptek, P., & Banks, M. (1978). Infant sensory assessment: Vision. In F. Minifie & L. Lloyd (Eds.), *Communicative and cognitive abilities: Early behavioral assessment* (pp. 61–106). Baltimore: University Park Press.

Schneider, W., & Bjorklund, D. (1998). Memory. In D. Kuhn & R. Siegler (Eds.), *Handbook of child psychology: Vol. 2. Cognition, perception, and language* (5th ed., pp. 467–521). New York: Wiley.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science, 237,* 1317–1323.

Shultz, T. (2001). Connectionist models of development. In N. J. Smelser & P. B. Baltes (Eds.), *International encyclopedia of the social and behavioral science* (Vol. 4, pp. 2577–2580). Oxford, England: Pergamon.

Shultz, T., Schmidt, W., Buckingham, D., & Mareschal, D. (1995). Modeling cognitive development with a generative connectionist algorithm. In T. Simon & G. Halford (Eds.), *Developing cognitive competence: New approaches to process modeling* (pp. 205–261). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Sirois, S., & Shultz, T. (1998). Neural network modeling of developmental effects in discrimination shifts. *Journal of Experimental Child Psychology, 71,* 235–274.

Sloman, S. A. (1997). Explanatory coherence and the induction of properties. *Thinking & Reasoning, 3,* 81–110.

Thomas, M., & Karmiloff-Smith, A. (in press). Connectionist models of development, developmental disorders and individual differences. In R. Sternberg, J. Lautrey, & T. Lubart (Eds.), *Models of intelligence for the next millennium.* Washington, DC: American Psychological Association.

Turkewitz, G., & Kenny, P. A. (1982). Limitations on input as a basis of neural organization and perceptual development: A preliminary theoretical statement. *Developmental Psychobiology, 15,* 357–368.

Turkewitz, G., & Kenny, P. A. (1985). The role of developmental limitations of sensory input on sensory/perceptual organization. *Developmental and Behavioral Pediatrics, 6,* 302–306.

Widrow, B., & Hoff, M. E. (1960). Adaptive switching circuits. In *IRE WESCON Convention Record* (pp. 96–104). New York: XXPUBLISHER NAME.

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

APPENDIX: THE MATHEMATICAL FORMULATION OF SUSTAIN

This appendix describes in detail the formalized operation of SUSTAIN, which is derived from its key psychological principles. We begin by describing our stimulus representation strategy. Following this we describe the mathematical equations that govern the model's operation. Finally, we discuss the parameters used in the simulations reported here.

Input Representation

Stimuli are represented in the model as vector frames where the dimensionality of the vector is equal to the dimensionality of the stimuli. The category label is also included as a stimulus dimension (the terms *dimension* and *attribute* are used interchangeably). Thus, stimuli that vary on three perceptual dimensions (e.g., size, shape, and color) and are members of one of two categories would require a vector frame with four dimensions. A four-dimensional binary-valued stimulus (three perceptual dimensions plus the category label) can be thought of as a four-character string (e.g., 1 2 1 1) in which each character represents the value of a stimulus dimension. For example, the first character could denote the size dimension, with a 1 indicating a small stimulus and a 2 indicating a large stimulus.

Of course, a learning trial usually involves an incomplete stimulus representation. For instance, in classification learning all the perceptual dimensions are known, but the category label dimension is unknown and queried. After the learner responds to the query, corrective feedback is provided. Assuming the fourth stimulus dimension is the category label dimension, the classification trial for this stimulus is represented as 1 2 1 ? → 1 2 1 1.

On every classification trial, the category label dimension is queried, and corrective feedback indicating the category membership of the stimulus is provided. In contrast, on inference learning trials, participants are given the category membership of the item but must infer an unknown stimulus dimension. Possible inference learning trials for the this stimulus description are ? 2 1 1 → 1 2 1 1, 1 ? 1 1 → 1 2 1 1, and 1 2 ? 1 → 1 2 1 1. Notice that inference and classification learning provide the learner with the same stimulus information after feedback (although the pattern of feedback varies).

Unsupervised learning does not involve informative feedback. In unsupervised learning, every item is considered to be a member of the same global category. Thus, the category label dimension is unitary valued and uninformative for differentiating between stimuli. However, the degree to which any particular stimulus activates this category dimension indicates the degree to which the network recognizes the stimulus.

To represent a nominal stimulus dimension that can display multiple values, SUSTAIN devotes multiple input units. To represent a nominal dimension containing *k* distinct values, *k* input units are utilized. All the units forming a dimension are set to 0, except for the one unit that denotes the nominal value of the dimension (this unit is set to 1). For example, the stimulus dimension of marital status has three values (single, married, divorced). The pattern [0 1 0] represents the dimension value of married. A complete stimulus is represented by the vector $I^{pos_{ik}}$ where *i* indexes the stimulus dimension and *k* indexes the nominal values for dimension *i*. For example, if marital status was the third stimulus dimension and the second value was present (i.e., married), then $I^{pos_{32}}$ would equal 1, whereas $I^{pos_{31}}$ and $I^{pos_{33}}$ would equal 0. The *pos* in I^{pos} denotes that the current stimulus is located at a particular position in a multidimensional representational space.

Receptive Fields

Each cluster has a receptive field for each stimulus dimension. A cluster's receptive field for a given dimension is centered at the cluster's position along that dimension. The position of a cluster within a dimension indicates the cluster's expectations for its members.

The tuning of a receptive field (as opposed to the position of a receptive field) determines how much attention is being devoted to the stimulus dimension. All the receptive fields for a stimulus dimension have the same tuning (i.e., attention is dimensionwide as opposed to cluster-specific). A receptive field's tuning changes as a result of learning. This change in receptive field tuning implements SUSTAIN's selective attention mechanism. Dimensions that are highly attended to develop peaked tunings, whereas dimensions that are not well attended to develop broad tunings. Dimensions that provide consistent information at the cluster level receive greater attention.

Mathematically, receptive fields have an exponential shape with a receptive field's response decreasing exponentially as distance from its center increases. The activation function for a dimension is:

$$\alpha(\mu) = \lambda e^{-\lambda\mu} \tag{1}$$

where λ is the tuning of the receptive field, μ is the distance of the stimulus from the center of the field, and $\alpha(\mu)$ denotes the response of the receptive field to a stimulus falling μ units from the center of the field. The choice of exponentially shaped receptive fields is motivated by Shepard's (1987) work on stimulus generalization.

Although receptive fields with different λ have different shapes (ranging from a broad to a peaked exponential), for any λ, the area "underneath" a receptive field is constant:

$$\int_0^\infty \alpha(\mu)d\mu = \int_0^\infty \lambda e^{-\lambda\mu}d\mu = 1. \tag{2}$$

For a given μ, the λ that maximizes α(μ) can be computed from the derivative:

$$\frac{\partial\alpha}{\partial\lambda} = e^{-\lambda\mu}(1-\lambda\mu). \tag{3}$$

These properties of exponentials prove useful in formulating SUSTAIN.

Cluster Activation

With nominal stimulus dimensions, the distance μ_{ij} (from 0 to 1) between the *i*th dimension of the stimulus and cluster *j*'s position along the *i*th dimension is:

$$\mu_{ij} = \frac{1}{2}\sum_{k=1}^{v_i}\left| I^{posik - H_j^{posik}} \right| \tag{4}$$

where v_i is the number of different nominal values on the *i*th dimension, *I* is the input representation (as described in a previous section), and $H_j^{pos_{ik}}$ is cluster *j*'s position on the *i*th dimension for value *k* (the sum of all *k* for a dimension is 1). The position of a cluster in a nominal dimension is actually a probability distribution that can be interpreted as the probability of displaying a value given that an item is a member of the cluster. For example, a cluster in which 20% of the members are single, 45% are married, and 35% are divorced will converge to the location [.20 .45 .35] within the marital status dimension. The distance μ_{ij} will always be between 0 and 1 (inclusive).

The activation of a cluster is given by:

$$H_j^{act} = \frac{\sum_{i=1}^{m}(\lambda_i)^r e^{-\lambda_i \mu_{ij}}}{\sum_{i=1}^{m}(\lambda_i)^r} \tag{5}$$

where H_j^{act} is the activation of the *j*th cluster, *m* is the number of stimulus dimensions, λ_i is the tuning of the receptive field for the *i*th input dimension, and *r* is an attentional parameter (always nonnegative). When *r* is large, input units with tighter tunings (units that seem relevant) dominate the activation function. Dimensions that are highly attended to have larger λs and will have greater importance in determining the clusters' activation values. Increasing *r* simply accentuates this effect. If *r* is set to 0, every dimension receives equal attention. Equation 5 sums the responses of the receptive fields for each input dimension and normalizes the sum (again, highly attended dimensions weigh heavily). Cluster activation is bound between 0 (exclusive) and 1 (inclusive). Unknown stimulus dimensions (e.g., the category label in a classification trial) are not included in this calculation.

Competition

Clusters compete to respond to input patterns and in turn inhibit one another. When many clusters are strongly activated, the output of the winning cluster H_j^{out} is less:

For the winning H_j with the greatest H^{act},

$$H_j^{out} = \frac{(H_j^{act})^\beta}{\sum_{i=1}^{n}(H_i^{act})^\beta} H_j^{act} \tag{6}$$

For all other H_j,

$$H_j^{out} = 0.$$

where *n* is the number of clusters, and β is the lateral inhibition parameter (always nonnegative) that regulates cluster competition. When β is small, competing clusters strongly inhibit the winner. When β is large, the winner is weakly inhibited. Clusters other than the winner have their output set to zero. Equation 6 is a straightforward method for implementing lateral inhibition. It is a high-level description of an iterative process where units send signals to each other across inhibitory connections. Psychologically, Equation 6 signifies that competing alternatives will reduce confidence in a choice (reflected in a lower output value).

Response

Activation is spread from the clusters to the output units of the queried (the unknown) stimulus dimension *z:*

$$C_{zk}^{out} = \sum_{j=1}^{n} w_{j,zk} H_j^{out} \tag{7}$$

where C_{zk}^{out} is the output of the output unit representing the *k*th nominal value of the queried (unknown) *z*th dimension, *n* is the number of clusters, and $w_{j,zk}$ is the weight from cluster *j* to category unit C_{zk}. A winning cluster (especially one that did not have many competitors and is similar to the current input pattern) that has a large positive connection to a output unit will strongly activate the output unit. The summation in the preceding calculation is not really necessary given that only the winning cluster has a nonzero output but is included to make the similarities between SUSTAIN and other models more apparent.

The probability of making response *k* (the *k*th nominal value) for the queried dimension *z* is:

$$Pr(k) = \frac{e^{\left(d \cdot C_{zk}^{out}\right)}}{\sum_{j=1}^{v_z} e^{\left(d \cdot C_{zj}^{out}\right)}} \tag{8}$$

where *d* is a response parameter (always nonnegative), and v_z is the number of nominal units (and hence output units) forming the queried dimension *z*. When *d* is high, accuracy is stressed and the output unit with the largest output is almost always chosen. The Luce (1959) choice rule is conceptually related to this decision rule.

Learning

After responding, feedback is provided to SUSTAIN. The target value for the *k*th category unit of the queried dimension *z* is:

$$t_{zk} = \begin{Bmatrix} max(C_{zk}^{out}, 1), \text{ if } I^{pos_{zk}} \text{ equals } 1. \\ min(C_{zk}^{out}, 0), \text{ if } I^{pos_{zk}} \text{ equals } 0. \end{Bmatrix} \tag{9}$$

Kruschke (1992) referred to this kind of teaching signal as a "humble teacher" and explained when its use is appropriate. Basically, the model is not penalized for predicting the correct response more strongly than is necessary.

A new cluster is recruited if the winning cluster predicts an incorrect response. In the case of a supervised learning situation, a cluster is recruited according to the following procedure:

For the queried dimension *z*,
if t_{zk} does not equal 1 for the C_{zk}
with the largest output C_{zk}^{out} of all C_{z*},
then recruit a new cluster. (10)

In other words, the output unit representing the correct nominal value must be the most activated of all the output units forming the queried stimulus dimension.

In the case of an unsupervised learning situation, SUSTAIN is self-supervising and recruits a cluster when the most activated cluster H_j's activation is below the threshold τ:

$$\text{if } (H_j^{act} < \tau)\text{, then recruit a new cluster.} \tag{11}$$

Unsupervised recruitment in SUSTAIN bears a strong resemblance to recruitment in adaptive resonance theory (Carpenter & Grossberg, 1987), Clapper and Bower's (1991) qualitative model, and Hartigan's (1975) leader algorithm.

When a new cluster is recruited, it is centered on the misclassified input pattern, and the clusters' activations and outputs are recalculated. The new cluster then becomes the winner because it will be the most highly activated cluster (it is centered on the current input pattern—all μ_{ij} will be 0). Again, SUSTAIN begins with a cluster centered on the first stimulus item.

The position of the winner is adjusted:

For the winning H_j,

$$\Delta H_j^{pos_{ik}} = \eta(I^{pos_{ik}} - H_j^{pos_{ik}}) \tag{12}$$

where η is the learning rate. The centers of the winner's receptive fields move toward the input pattern according to the Kohonen (1989) learning rule. This learning rule centers the cluster amidst its members.

Using our result from Equation 3, receptive field tunings are updated according to:

$$\Delta\lambda_i = \eta e^{-\lambda_i \mu_{ij}}(1 - \lambda_i \mu_{ij}) \tag{13}$$

where j is the index of the winning cluster.

Only the winning cluster updates the value of λ_i. Equation 13 adjusts the peakedness of the receptive field for each input so that each input dimension can maximize its influence on the clusters. Initially, λ_i is set to be broadly tuned with a value of 1. The value of 1 is chosen because the maximal distance μ_{ij} is 1 and the optimal setting of λ_i for this case is 1 (i.e., Equation 13 equals 0). Under this scheme, λ_i cannot become less than 1, but can become more narrowly tuned.

When a cluster is recruited, weights from the unit to the output units are set to 0. The one layer delta learning rule (Widrow & Hoff, 1960) is used to adjust these weights:

$$\Delta w_{j,zk} = \eta(t_{zk} - C_{zk}^{out})H_j^{out}. \tag{14}$$

where z is the queried dimension. Note that only the winning cluster will have its weights adjusted because it is the only cluster with a nonzero output.

SUSTAIN Parameters

The parameters used in the simulations reported in this study are shown in Table 1. The learning rate, cluster competition, decision consistency, and attentional focus parameters were fixed from the best fitting parameter for previously published studies. Thus, only threshold and blur were free parameters for the reported studies.