Journal of Biological Systems, Vol. 2, No. 3 (1994) 283–305

ANALYSIS OF CATEGORICAL TIME-SERIES TEXT RECALL DATA USING A CONNECTIONIST MODEL

R. M. GOLDEN
School of Human Development, The University of Texas at Dallas, GR. 41, Box 830688, Richardson, Texas 75083-0688, USA

ABSTRACT

Categorical time-series are generated by discrete-time probabilistic dynamical systems which can only be in one of a small number of finite states at any given instant in time. A novel statistical methodology based upon log-linear modelling is proposed for analyzing categorical time-series data which allows one to incorporate a considerable amount of prior knowledge directly into the data analysis. The statistical model can be shown to be formally equivalent to a connectionist (i.e., artificial neural network) model. Methods for model selection and hypothesis testing using the new statistical model for samples with large numbers of observations are then developed using asymptotic statistical theory. To illustrate this new method of categorical time-series data analysis, the model is applied to the analysis of text free recall data from children and adults. These analyses indicated that the model can successfully use the order of recalled text propositions to discriminate among alternative theories of prior knowledge and alternative treatment conditions. The reliability of the large sample statistical tests were also checked using a boot-strap methodology and found to be acceptable.

Keywords: Connectionist, asymptotic statistics, text memory recall, causal network, semantic network, misspecification.

1. Categorical Time-Series Modelling

Consider a complex system which can enter into only one of the d states $\{\mathbf{f}_1, \ldots, \mathbf{f}_d\}$ at each instant in time where the feature vector $\mathbf{f}_i$ is a d-dimensional vector whose ith element is set equal to one and whose remaining $d-1$ elements are set equal to zero. For example, imagine a simple organism which can either: (i) sleep ($\mathbf{f}_1$), (ii) eat ($\mathbf{f}_2$), or (iii) move ($\mathbf{f}_3$). In this case $d = 3$. At regular time intervals, a scientist records the organism's state as belonging to one of the three possible feature categories. Thus, the data for this experiment might have a format given by the following ordered set such as:

$$\{\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_2, \mathbf{f}_1, \mathbf{f}_3, \mathbf{f}_3, \mathbf{f}_3, \mathbf{f}_2\}\,. \tag{1.1}$$

The sequence of feature vectors in (1.1) is referred to as a *trajectory* whose *length* is equal to 8. The length of a trajectory may vary randomly as well. This is an example of the type of data which is suitable for *categorical time-series data analysis.*

Categorical time-series data analyses are applicable to many areas of cognitive science and experimental psychology. For example, Allison and Liker [1] have used categorical time-series models to study social interactions between individuals. As another example, hidden Markov models of speech recognition [27] are based upon the assumption of a categorical time-series as well. A third example of categorical time-series data analysis which will be discussed in greater detail in this article is applicable to the analysis of text recall data. In some cases, a text may be represented as an ordered sequence of propositions or text features. The recall protocol of a subject can then be viewed as a trajectory whose feature categories correspond to the text features of the original story [13,14].

Golden [10,8,9] and White [45] have noted the close relationship between connectionist and statistical modelling. In this paper, a novel method for the analysis of categorical time-series data is proposed using a statistical model which has a connectionist interpretation. The analysis begins with a *feature coding table* which consists of d features where $\mathbf{f}_j$ indicates the identity of feature j. The data generating process generates sample trajectories of the form: $\mathbf{x}(1), \mathbf{x}(2), \ldots, \mathbf{x}(M)$ where M is a random variable and $\mathbf{x}(t)$ is a random vector which can take on any one of the d values specified in the feature coding table. Thus, $\mathbf{x}(t) \in \{\mathbf{f}_1, \ldots \mathbf{f}_d\}$. As in all statistical analyses, two types of problems are addressed. The "model selection problem" where the most likely model of the data generating process is chosen, and the "hypothesis testing problem" where the model is used as a tool for examining effects of treatment condition.

It should be emphasized again that the application of the proposed connectionist model to the analysis of text recall data should only be considered one of many possible potential applications. The proposed method of constrained categorical time-series analysis should be useful to researchers in many different fields who are interested in uncovering and exploiting temporal regularities in stochastic dynamical systems whose states can take on only one of a small number of permissible values.

1.1. *Relationships with Existing Data Analysis Methods*

Perhaps the most standard statistical tool for the analysis of categorical time-series data is the log-linear model (see [2,16,17] for reviews and see for example SAS program CATMOD [23]), which is a method of analysis analogous to standard linear analysis of variance. The categorical time-series analysis model proposed in this article is also a log-linear model but differs from the standard theory in at least three fundamental ways. First, if three trajectories each of length M are observed, then the standard log-linear theory associated with a K-order Markov process would associate roughly $3M/(K+1)$ data points with the three trajectories. The model proposed in this article considers each trajectory to be a single data point. Thus, according to the model proposed here, the number of data points in this example is equal to three. For some types of categorical time-series analyses, such a conservative approach for counting data points is more appropriate than the

standard log-linear theory. For example, most experiments in the field of psychology observe the behavior of several subjects in several experimental conditions over time. Such a situation corresponds to the case of multiple trajectories with definite starting points and fixed (or variable) stopping points.

Second, the proposed categorical time-series data analysis is based upon a ***highly constrained*** parametric log-linear model unlike the standard log-linear theory. Thus, the proposed categorical time-series data analysis may be viewed as a specialized type of path analysis [2]. Construction of a suitably constrained log-linear model allows the incorporation of a considerable amount of prior knowledge directly into the data analysis. In particular, the proposed analysis is especially useful for situations where: (i) the amount of time-series data is relatively small, (ii) the dimensionality d of the feature table is relatively large and (iii) some prior knowledge is available regarding how the likelihood of one feature in the system is functionally dependent upon the other features. One reason why log-linear categorical time-series models are not typically used in the literature is because they require an excessive number of data points. Incorporation of prior knowledge of the data generating process into the statistical model can severely reduce the number of free parameters in a categorical log-linear model resulting in more effective inferences with less data. Of course, one must be careful when introducing such information because the resulting data analyses are only interpretable with respect to the assumed prior knowledge framework.

The third important difference between the proposed method of categorical time-series data analysis and popular methods is motivated by the second difference which assumes a highly constrained model. If a statistical model is highly constrained, it is quite likely that the model will not be correctly specified. Accordingly, the methods for model selection and hypothesis-testing presented here are derived under the assumption that the proposed model is not necessarily correctly specified (i.e., the proposed model may not be capable of adequately modelling the data generating process). Such an approach based upon the methods of [44] and [43] and reviewed by [11] guarantees that decisions made in the course of model selection and hypothesis testing will be correct in the presence of many common forms of model misspecification. For example, most researchers currently use Wilks [47] generalized likelihood ratio test as a basis for model selection and hypothesis testing but this statistical test is known to yield incorrect decisions when the full model is not correctly specified (e.g., [44,46,43,11]).

1.2. *Overview*

One exciting new area of application of the proposed constrained categorical time-series analysis model is the analysis of text recall data [12]. In this approach, the feature table consists of the d text features or propositions that are used to code a subject's recall of a particular text. Most current models of text recall such as the Kintsch and Van Dijk [25] or Fletcher and Bloom [6] models can only predict the

probability that a particular text feature is included in a text. The proposed constrained categorical time-series analysis model explicitly uses the temporal structure of free recall data as a dependent measure.

This paper is organized as follows: First a connectionist model of text recall is described. This model is formally equivalent to the proposed constrained categorical time-series analysis model. Second, procedures for using the model to examine effects of different treatment conditions are described. And third, methods for deciding which of several versions of the model "best fits" a given set of recall data are described.

2. Description of the Model

Although the model is technically a very specific type of structured path analysis designed for categorical (nominal) dependent variables, the model will be presented for expository reasons as a connectionist model. The connectionist model interpretation provides a good intuitive motivation for the specific assumptions of the underlying statistical model.

2.1. *Representational Assumptions*

A feature coding table specifies a particular *situation state space* [12,13,11,15,14] which is a d-dimensional space whose elements are feature vectors $\mathbf{f}_1, \ldots, \mathbf{f}_d$. The feature vectors have a local representation so that only one of the d text features can be active at any moment of time. In particular, let $\mathbf{x} = [x_1, \ldots, x_d]^T$ be a point in the d-dimensional situation state space such that $\mathbf{x}$ is constrained to take on only one of d values: $\mathbf{f}_1, \ldots, \mathbf{f}_d$ where $\mathbf{f}_j$ is a vector of zeros with a 1 in the jth position.

As an example, consider the situation state space for a simplified version of the story *Jack and Jill* defined by the following feature table with $d = 3$.

Table 1

Text Feature Category	Sentence Fragment of Original Text
$\mathbf{f}_1$: Go (Jack and Jill, To Hill)	Jack and Jill went up a hill.
$\mathbf{f}_2$: Want (Jack and Jill, Water)	Jack and Jill wanted a drink of water.
$\mathbf{f}_3$: Drink (Jack,Water)	Jack drank some water.

The story *Jack and Jill* is modelled using the above feature table as the ordered set of text features: $\{\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3\}$. Table 1 is also used for the coding of recall data. For example, suppose a subject recalls the story *Jack and Jill* as follows: *Jack and Jill wanted a drink of water so they went to a hill.* This recall protocol would be coded as the ordered set of text features: $\{\mathbf{f}_2, \mathbf{f}_1\}$.

A subset of semantic relationships of the same type which share the same relative influences upon the model's behavior and are referred to as a *knowledge schema* which is formally represented as a directed graph or digraph. More formally, let the nominal variable R indicates the *type* of a particular knowledge schema. For

example, one value of R might indicate that the knowledge schema specifies *causal* relationships while another value of R might indicate the knowledge schema specifies *syntactic* or *referential* relationships. Let the d-dimensional matrix $\mathbf{W}$ represent a knowledge schema of type R with respect to a feature table consisting of d features. A *relationship* is an ordered pair $(\mathbf{f}_i, \mathbf{f}_j)$ of text features defined with respect to a knowledge schema of type R. This relationship is represented by setting the element, w_{ij} in the digraph matrix $\mathbf{W}$ equal to some non-zero value. The *relative strength* of the relationship $(\mathbf{f}_i, \mathbf{f}_j)$ with respect to the knowledge schema of type R is given by the value w_{ij}. Currently, $w_{ij} = 1$ indicates the presence of relationship $(\mathbf{f}_i, \mathbf{f}_j)$ and $w_{ij} = 0$ indicates the absence of relationship $(\mathbf{f}_i, \mathbf{f}_j)$.

A very important type of knowledge schema is a *semantic* knowledge schema. Any type of semantic relationship can be expressed in this theoretical framework. Thus, semantic relationships for representing *referential coherence* [25] or generic knowledge structure (GKS) representation theory [19] can be modelled within this framework. Currently, three types of semantic knowledge schemata are implemented in the model. The first is a *pure episodic knowledge schema*, $\mathbf{W}^{(1)}$, whose pair-wise relationships indicate when one text feature $\mathbf{f}_i$ immediately follows another text feature $\mathbf{f}_j$ (but $\mathbf{f}_j$ does not cause $\mathbf{f}_i$ to occur). The second is a *pure causal knowledge schema*, $\mathbf{W}^{(2)}$, whose pair-wise relationships indicate text feature $\mathbf{f}_j$ causes text feature $\mathbf{f}_i$ to occur according to Mackie's [29] counterfactual notion of causality but $\mathbf{f}_j$ does not immediately precede $\mathbf{f}_i$ in the text. The third is a *shared episodic-causal knowledge schema*, $\mathbf{W}^{(3)}$, whose pair-wise relationships indicate: (i) text feature $\mathbf{f}_j$ immediately precedes text feature $\mathbf{f}_i$, and (ii) $\mathbf{f}_j$ causes $\mathbf{f}_i$ to occur according to a counterfactual notion of causality.

2.2. *Statistical Regularity Modelling Assumptions*

2.2.1. *Working Memory Modelling Assumption*

The activation pattern $\mathbf{y}(t)$ in Fig. 1 over the model's input units indicates the contents of the model's *working memory buffer*. More formally, the model's *working memory buffer* is updated according to the formula:

$$\mathbf{y}(t) = \mathbf{x}(t-1) + \mathbf{Q}\mathbf{y}(t-1)\,. \tag{2.1}$$

The constant *local coherence strategy* matrix $\mathbf{Q}$ can be chosen to model versions of various types of psychologically relevant local coherence strategies such as the recency [25] leading edge [25] or current state [6] strategies.

For example, a version of the *recency encoding strategy* where the reader keeps the most recently activated text features in a limited capacity working memory in order to maintain coherence during text comprehension is modelled by setting the matrix $\mathbf{Q} = \mu\mathbf{I}$ where $0 < \mu < 1$ and $\mathbf{I}$ is the identity matrix. Moreover, a version of the *leading edge encoding strategy* can be implemented by choosing $\mathbf{Q}$ such that: (i) the recency strategy is implemented, and (ii) previously activated superordinate nodes in working memory $\mathbf{y}(t-1)$ inhibit their currently active subordinate nodes in

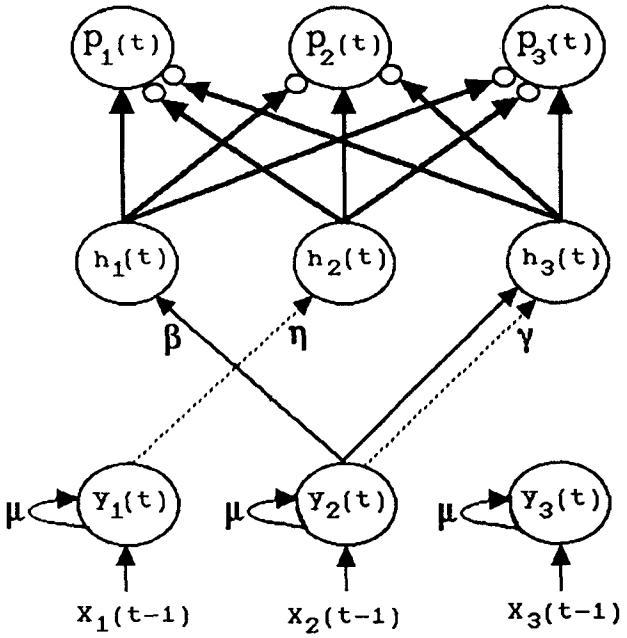

Fig. 1. A connectionist representation of the categorical time-series analysis statistical model. The notation $x_i(t-1) = 1$ indicates that text feature i was the $(t-1)$th feature recalled by the subject. The notation $y_i(t)$ indicates that text feature i is active in the model's working memory at time t. The probability $p_i(t)$ is the model's expectation that text feature i will be the tth feature recalled by the subject. In this simple example, the episodic digraph contains exactly one link (single dashed line) between text feature 1 and text feature 2. The causal digraph consists of exactly one link (single solid line) between text feature 2 and text feature 1. The shared causal-episodic digraph consists of exactly one link (dashed and solid line pair) between text feature 2 and text feature 3.

working memory $\mathbf{y}(t)$. And finally, in a manner similar to the leading edge strategy, a version of Fletcher and Bloom's [6] *current state local coherence strategy* may be modelled by picking $\mathbf{Q}$ such that: (i) the recency strategy is implemented and (ii) previously activated causal consequence nodes in working memory $\mathbf{y}(t-1)$ inhibit their currently activated causal antecedent nodes in working memory $\mathbf{y}(t)$.

2.2.2. *Superposition of Knowledge Schemata Assumption*

The connections from the working memory buffer activation pattern

$$\mathbf{y}(t) = [y_1(t), \ldots, y_d(t)] \tag{2.2}$$

to the hidden unit layer activation pattern $\mathbf{h}(t) = [h_1(t), \ldots, h_d(t)]$ are now considered. Each set of connections corresponds to a semantic knowledge schema digraph such as the *pure episodic*, *pure causal* or *shared episodic-causal* schemata described earlier. The jth semantic digraph or knowledge schema is represented by the matrix $\mathbf{W}^{(j)}$.

Specifically, this assumption is instantiated by the following equation:

$$\mathbf{h}(t) = [\eta \mathbf{W}^{(1)} + \beta \mathbf{W}^{(2)} + \gamma \mathbf{W}^{(3)}]\mathbf{y}(t) \tag{2.3}$$

where η is the trace strength of the episodic digraph $\mathbf{W}^{(1)}$, β is the trace strength of the causal digraph $\mathbf{W}^{(2)}$ and γ is the trace strength of the shared causal-episodic digraph $\mathbf{W}^{(3)}$. Thus, only three free parameters completely specify the pattern of connection strengths in the network. Note that additional digraphs can be easily introduced at the cost of adding additional free parameters to the model.

Evidence that increasing the number of causal links to a proposition in a text increases the likelihood that the proposition will be recalled [18,20,39,40] is consistent with the superposition assumption when a causal digraph is involved. The unique estimates of η, β and γ are numerically estimated using the quasi-maximum likelihood estimation procedure described by Golden [11].

2.2.3. *Response Competition Modelling Assumption*

The connections from the hidden layer $\mathbf{h}(t) = [h_1(t), \ldots, h_d(t)]$ to the output unit layer $\mathbf{p}(t) = [p_1(t), \ldots, p_d(t)]$ are fixed and not modifiable. These connections are chosen so that the ith output unit activation may be interpreted as the probability that text feature i is the next feature in the recall protocol. This "response competition" assumption is formally instantiated as:

$$p_i(t) = p_i[x_i(t+1) = 1 | y(t)] = \frac{\exp\{h_i(t)\}}{\sum_{j=1}^{d} \exp\{h_j(t)\}}, \tag{2.4}$$

where $h_i(t)$ is the activation of the ith hidden unit and $p_i(t)$ is the activation of the ith output unit given the working memory buffer contents $y(t)$. Note that probabilistic choice rules of the general form of (2.4) have been used successfully by a number of memory theorists [7,24,28,33,37]. In particular, the sampling assumptions of (2.4) are very similar to the SAM model and were used by SAM to explain the *list length* effect [7,33,37] in free recall. In addition, the response competition assumption is also consistent with experimental evidence that people make elaborative inferences only in predictive contexts [32,38,41].

2.2.4. *Trajectory Generation Modelling Assumption*

The number of items included in a recall protocol is modelled as a Poisson random variable with mean λ. The *expected trajectory length* context parameter λ is both task-dependent and text-dependent. The Poisson modelling assumption essentially means that a subject's decision to stop generating the ith item is totally independent of the contents of the subject's generated recall protocol at that instant in time. The Poisson modelling assumption is reasonable for modelling factors such as motivation, experimental time-constraints or text length. As noted by [11], the parameter λ may be estimated according to a simple closed form expression.

3. The Model Selection Problem

3.1. *Theoretical Background*

The first stage of the constrained categorical time-series analysis is referred to as the "model selection problem." In this stage, a relatively small number of alternative versions of the model are considered. The problem is to choose the model which "best-fits" the set of recall data. This problem is solved using the following three step procedure. First, estimate the parameters of each of the alternative versions of the model with respect to the same data set. Second, compute the likelihood of each alternative model with respect to the same set of recall data. And third, choose the model which makes the data set most probable.

Such a strategy is also formally equivalent to computing the KLIC (Kullback-Leibler Information Criterion) error measure for each model, and choosing the model with the smallest KLIC error (see [11,43,44] for reviews of this approach). For example, the popular Wilks [47] generalized likelihood ratio test is naturally derived from the strategy of choosing the model which makes the observed data most likely. An explicit formula for computing the KLIC error for the model described in this article is provided in [11]. Note that the KLIC error is also referred to as the "divergence" measure or cross-entropy [26] between the parameterized probability model and an estimate of the probability distribution which originally generated the data.

In many cases, a statistical test for deciding if the KLIC errors for two different models are significantly different from one another is desirable. Vuong [43] has showed how to construct such a test. Golden [11] applied Vuong's [43] general theory to the log-linear categorical time-series model discussed here in order to derive appropriate statistical tests for model selection in the presence of model misspecification. Golden's [11] application of Vuong's [43] theory to the categorical statistical model described previously will now be briefly reviewed.

For each recall protocol (i.e., trajectory), the KLIC error for that protocol is computed under Model 1 and then is computed under Model 2. The difference between the KLIC error for Model 1 *minus* the KLIC error for Model 2 for the jth protocol associated with experimental condition i is defined as δ_j^i. The setup for these calculations is shown in Table 2 using a simplified example involving only three recall protocols collected in experimental condition i.

Table 2

	Model 1	Model 2	δ^i
Protocol 1	4	3	$\delta_1^i = 4 - 3 = 1$
Protocol 2	7	0	$\delta_2^i = 7 - 0 = 7$
Protocol 3	3	2	$\delta_3^i = 3 - 2 = 1$

The quantity:

$$\bar{D}_i = (1/N) \sum_{j=1}^{N} \delta_j^i \tag{3.1}$$

converges to the population mean Δ^i which can be shown to be the true difference between the population KLIC errors for the two models for expected condition i. If $\Delta^i > 0$, then model 2 fits the data collected from condition i better than model 1. $\Delta^i < 0$, then model 1 fits the data collected from condition i better than model 2. If $\Delta^i = 0$, the two models fit the data collected from condition i equally effectively.

By the central limit theorem [43], for large sample sizes $\bar{D}_i$ has a Gaussian distribution with estimated mean $\bar{D}_i$ and estimated variance:

$$s^2_{D_i} = \sum_{j=1}^{N} (\delta^i_j - \bar{D}_i)^2 / N^2 . \tag{3.2}$$

Thus, a statistical test designed to test the null hypothesis, $H_o : \Delta^i = 0$ may be constructed by computing:

$$V^i_{obs} = \frac{\bar{D}_i}{s_{D_i}} . \tag{3.3}$$

Let a scalar Z_{crit} be chosen such that: The probability that the absolute value of a Gaussian random variable with mean zero and variance 1 exceeds Z_{crit} is equal to the significance level, α. If $V_{obs} > Z_{crit}$, then Model 2 is better than Model 1. If $V_{obs} < -Z_{crit}$, then Model 1 is better than Model 2. If $|V_{obs}| < Z_{crit}$, then the models are "equally distant" from the environmental distribution. Note the similarity of this analysis to the classical "within-groups" t-test which is frequently used in the fields of biology and psychology (e.g., [22]). Finally, it should be noted that an additional statistical test is required to test the null hypothesis that s^2_D does not converge to zero as the number of observations becomes large. If s^2_D does indeed converge to zero for large N, then it turns out that the two models should be considered to be equally distant from the *environmental distribution* which originally generated the data [43].

Vuong's analysis has two important advantages over classical generalized likelihood ratio testing. First, classical generalized likelihood ratio testing using Wilk's [47] approach requires that at least one of the models adequately "fits the data" This assumption is not required for the statistical analysis proposed here. Second, classical generalized likelihood ratio testing using Wilk's [47] approach requires that one of the models must be generated by constraining a subset of parameters of the "full" model to be equal to constant values. This "fully nested" assumption is also not required for the statistical analysis proposed here.

These procedures may be used to make comparisons of alternative models of text recall. Two alternative models of the recall data generated by subjects recalling the same story are first constructed. The first model, for example, could be derived from a causal digraph generated by an "expert" analysis of the text. The second model could be derived from a different causal digraph derived empirically from human subject rating data where human subjects are asked to rate the degree of causal relatedness between all pairs of text features. Then, as previously noted, the

estimated mean KLIC error difference between the two models is computed and checked to see if that estimated difference is significantly different from zero.

Usually, the researcher will wish to attempt to draw conclusions about model variations across experimental conditions. For example, the researcher may want to test the null hypothesis that one method of causal digraph construction is superior to another method regardless of which text is being analyzed. Let $\bar{D}_i$ and s_{D_i} be the mean KLIC error difference and its standard error for the ith experimental condition. The KLIC error difference between the two models ***across k conditions*** is then defined as:

$$\bar{D}_c = (1/k) \sum_{i=1}^{k} \bar{D}_i \tag{3.4}$$

which has standard error given by:

$$s_{D_c} = (1/k) \left[\sum_{i=1}^{k} s_{D_i}^2 \right]^{1/2} . \tag{3.5}$$

Thus, to compare the two methods of digraph construction across k conditions, the statistic:

$$V_c = \frac{\bar{D}_c}{s_{D_c}} \tag{3.6}$$

is constructed. Let Z_{crit} be chosen such that the absolute value of a Gaussian random variable with mean zero and variance one is greater than Z_{crit} with probability α. Then, if the statistic $|V_c| > Z_{crit}$, the two methods of digraph construction are significantly different across the k experimental conditions. Moreover, the better method of digraph construction is the method with the smaller KLIC error.

3.2. *Applications of the Model Selection Test*

The model selection test described in the previous section was used to analyze the recall data collected by Golden and his colleagues [12]. In the ***adult recall study***, four stories analyzed by Trabasso *et al.* [39] were presented to 24 college students. Each college student read and recalled two of the same four stories in both an immediate recall condition and a one-week delayed recall condition. Each college student was instructed to recall the story they had previously read in the order in which the story had been originally presented. Subjects wrote their recall of each story using the third statement of the story as an initial retrieval cue. A distractor task consisting of arithmetic problems was presented to subjects in the immediate recall condition before they recalled their two stories from memory. In the ***child recall study***, two fifth grade/sixth grade classes of children each individually read one of the four stories, and then individually wrote their recall of the stories from memory after an intervening arithmetic distractor task. In total, the data set consists of twelve experimental conditions since each of the four stories were recalled in three distinct

conditions (children recall, adult immediate recall, and adult one week delayed recall).

3.2.1. *Finding the Best-Fitting Local Coherence Strategy*

The model was then fit to the recall data gathered from both the adult and child recall studies using five different values $(0, 1/8, 2/8, 3/8, 4/8)$ for the working memory span parameter μ using the recency encoding strategy $\mathbf{Q} = \mu\mathbf{I}$ where $\mathbf{Q}$ is defined as in (2.1) and $\mathbf{I}$ is the identity matrix. The version of the model with the smallest KLIC error was the model which used $\mu = 2/8$ so this value for μ was used in subsequent data analyses. Vuong's [43] (also see [11]) model selection tests established in post-hoc data analyses that the model was relatively insensitive to changes in the constant μ across experimental conditions. Based upon additional post-hoc data analyses, we are currently considering the possibility of making the "time-window" constant μ a text-dependent constant.

3.2.2. *Evidence for the Importance of Enabling Causal Links*

Consider the following notion of causality. Suppose that event 1 causes event 2 if event 1 is provided by subjects as the reason *why* event 2 occurred in a question-answering task. Graesser and his colleagues [20,18] have shown that causal digraphs derived from this notion of causality could be used to predict which statements in a story would be most likely to be recalled from memory.

On the other hand, Trabasso, Van Den Broek and their colleagues [39,40] have analyzed stories using a different definition of causality. In particular, Trabasso, Van Den Broek and their colleagues have found that a *counterfactual* (i.e., enabling) notion of causality is effective for predicting which statements in a text will be recalled from memory. In particular, the counterfactual notion of causality [29] states that event 1 causes event 2 provided that: If event 1 had not occurred in the circumstances, then event 2 would not have occurred.

In order to compare these two different notions of causality, a *causal relatedness rating study* involving University of Texas at Dallas adult subjects was done. Subjects were presented all possible pairs of propositions (text features) from all four stories. Each subject was asked to indicate on a scale of one to four the degree of causal relatedness between each pair of propositions. The subjects were told that:

> one indicated the absence of a causal relationship, two indicated a weak causal relationship, three indicated a moderately strong causal relationship and four indicated a strong causal relationship.

Using these rating data, a causal digraph was constructed such that each link had an average value of two or greater. This causal digraph was defined to be the *enabling* or *weak* causal digraph. A subset of this causal digraph was also generated such that each link had an average value of three or greater. This second causal digraph was defined as the *strong* causal digraph.

Two versions of the model were then constructed. One version used the strong causal digraph, while the other version used the weak causal digraph. The four free parameters were first estimated for the first model which used the strong causal digraph, and then were estimated for the second model which used the weak causal digraph. The KLIC error for the model with the strong causal digraph ($\hat{E} = 23.31$) was greater than the KLIC error for the model with the weak causal digraph ($\hat{E} = 22.43$). Moreover, this difference was statistically significant ($V_{obs} = -2.27, p < 0.05$) indicating that the *enabling* connections in the weak causal digraph were useful in explaining the data. The reliability of the statistical analysis was assessed by using a bootstrap analysis which yielded the same results. Currently, I am attempting to replicate this finding using a much larger database of stories and improved methods of causal digraph construction. Nevertheless, the current findings indicate that sequential regularities for model discriminations of this type are indeed present in the recall protocol data and that the model is sensitive enough to detect those temporal regularities.

3.2.3. *Evidence for the Importance of Unidirectional Causal Links*

Another question of interest was the bidirectionality of the links in the causal digraph model. Some experimental evidence supports the hypothesis that the probability of recalling a causal consequent given its antecedent is roughly the same as the probability of recalling a causal antecedent given its consequent ([3,36]; also see [34,35] for similar findings involving paired-associate word lists). On the other hand, using more sensitive response time measures, subjects process a causal consequent preceded by its causal antecedent more rapidly than a causal antecedent preceded by its causal consequent [21,42].

One interpretation of the above findings might be that the underlying causal associations are not bidirectional but standard data analysis methods are not sufficiently sensitive to pick up effects of input order for causally related statements. To explore this hypothesis further, the weak causal digraph model developed in the previous investigation was used to generate a *bidirectional* causal digraph. That is, if a semantic causal relationship $(\mathbf{f}_i, \mathbf{f}_j)$ was present in the original weak causal model, then *both* of the semantic causal relationships $(\mathbf{f}_i, \mathbf{f}_j)$ and $(\mathbf{f}_j, \mathbf{f}_i)$ were included in the bidirectional causal model. For purposes of experimental control, an additional *backward unidirectional* causal digraph model was constructed such that for each causal relationship $(\mathbf{f}_i, \mathbf{f}_j)$ in the original weak causal model, a backward causal relationship $(\mathbf{f}_j, \mathbf{f}_i)$ was included in the backward causal model.

Analysis of the recall data indicated that the directionality of causal semantic relationships was important. The original weak causal model KLIC error ($\hat{E} = 22.43$) was less than the error for the bidirectional model ($\hat{E} = 23.17$). Moreover, this error difference was statistically significant ($V_{obs} = -2.26, p < 0.05$). The original weak causal model KLIC error ($\hat{E} = 22.43$) was also significantly less than the error for the backward causal model ($\hat{E} = 24.11; V_{obs} = -3.4, p < 0.001$). These findings, in

conjunction with the experimental findings cited previously, are consistent with the hypothesis that the proposed statistical model can uncover temporal regularities in free recall data which current methods of data analysis cannot detect. Additional studies using additional texts and better digraph construction techniques are currently being planned to replicate and confirm this preliminary finding.

4. The Hypothesis Testing Problem

The end result of the model selection problem is to choose the parametric model that best fits the data generating process. Once such a model is selected, different model parameters will generate recall data sequences with different types of statistical regularities. For example, if subjects read and then immediately recall a text, it is likely that their recall performance will be greatly influenced by their episodic memory of the original text. Thus, the episodic strength parameter of the model should be relatively large. On the other hand, if subjects read a text and are asked to recall the text from memory one week later, then recall performance will be less influenced by their episodic memory of the text. Thus, the episodic strength parameter of the model should be relatively small.

These considerations suggest the following approach to data analysis using the model. The parameters and their asymptotic variance are estimated for a set of text recall data generated by subjects in one treatment condition of the experiment. In addition, the parameters and their asymptotic variance are estimated for a second treatment condition. A statistical test can then be constructed to decide if the parameter estimates computed from the data set associated with treatment 1 are significantly different from the parameter estimates computed from the data set associated with treatment 2. The problems of deriving, using, and interpreting such statistical tests are considered to be special cases of the more general *hypothesis testing problem.*

4.1. *Theoretical Background*

The procedures for estimating model parameters and their asymptotic variance are described by White [44,45,46]. Golden [11] explicitly derived formulas for numerically estimating the parameters and their asymptotic variance using White's asymptotic statistical theory. Briefly, let ρ_j^i be the parameters of the model estimated for treatment condition i and text j. In the example considered in this article, the dimensionality of ρ_j^i is four where the first component of ρ_j^i is the episodic strength parameter (η_j^i), the second component is the causal strength parameter (β_j^i), the third component is the shared causal-episodic strength parameter (γ_j^i), and the fourth component is the average number of items recalled (λ_j^i). Thus, $\rho_j^i = [\eta_j^i \;\; \beta_j^i \;\; \gamma_j^i \;\; \lambda_j^i]$. The goal of the parameter estimation process is to obtain parameter estimates that converge to those model parameters that best fit the data. The distance between the model and the data is computed using the KLIC error measure [26,44]. White [44,45,46] showed that for sufficiently large data sets, the

distribution of ρ_j^i is Gaussian with mean ρ_{j*}^i and covariance matrix $\mathbf{Q}_{j*}^i$ where ρ_{j*}^i is the parameter vector that minimizes the KLIC distance and $\mathbf{Q}_{j*}^i$ can be estimated from the first and second derivatives of the KLIC distance function (see [11] for a review and computation of the necessary derivatives).

Now construct a column vector ρ of the form:

$$\rho = [\rho_1^1,\ \rho_2^1,\ \rho_3^1,\ \rho_4^1, \rho_1^2,\ \rho_2^2,\ \rho_3^2,\ \rho_4^2,\ \rho_1^3,\ \rho_2^3,\ \rho_3^3,\ \rho_4^3\ ]^T\,, \tag{4.1}$$

where for expository reasons we have assumed just three treatment conditions and four texts in each treatment condition. Because the recall data between texts and conditions is assumed to be statistically independent, the covariance matrix of ρ, $\mathbf{Q}$, is given by:

$$\mathbf{Q} = \begin{bmatrix} \mathbf{Q}_1^1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \mathbf{Q}_2^1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \mathbf{Q}_3^1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{Q}_4^1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathbf{Q}_1^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \mathbf{Q}_2^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{Q}_3^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{Q}_4^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{Q}_1^3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{Q}_2^3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{Q}_3^3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{Q}_4^3 \end{bmatrix}$$

Finally, consider the null hypothesis:

$$H_o : \mathbf{S}\rho = \mathbf{0}\,, \tag{4.2}$$

where $\mathbf{0}$ is a column vector of zeros with the same dimensionality as ρ, and the matrix $\mathbf{S}$ is called the *selection matrix* that can be used to set up various types of contrasts involving the ρ vector.

The Wald test (as recommended by White, [44]) states that the null hypothesis H_o in (4.2) is rejected if the Wald statistic, W_n, given by:

$$W_n = n\rho^T(\mathbf{S}^T\mathbf{Q}\mathbf{S})^{-1}\rho \tag{4.3}$$

is significantly different from zero. Note that n is the number of recall protocols for each condition of the experiment that were used to estimate ρ. It can be shown that W_n has a chi-squared distribution with degrees of freedom equal to the rank of the selection matrix $\mathbf{S}$. This is the basic hypothesis testing tool that will be exploited.

One additional comment regarding the reliability of White's asymptotic statistical theory should be made. The formulas for $\mathbf{Q}$ are based upon asymptotic approximations that are only valid for sufficiently large data sets (i.e., the approximations are only valid for sufficiently large n). For this reason it is prudent, if

possible, to check the calculation of **Q** by estimating the covariance matrix of ρ in a second way. This has been done for all simulations reported here using the boot-strap method as described by Efron [4,5]. The agreement between these two different asymptotic approximations for estimating **Q** is quite good [11].

4.2. *Application of Hypothesis Testing*

4.2.1. *Introduction and Motivation*

Golden and his colleagues [12] used the analyses described here to analyze text recall data from children and adults. The Golden *et al.* study involved four different texts in three different treatment conditions. The treatment conditions were: (i) children's recall of the four texts, (ii) adults' immediate recall of the four texts and (iii) adults' recall of the four texts after a one-week delay. Golden and colleagues were interested in considering the following experimental questions. First, are there effects of the causal structure parameter (β)? Second, as retention interval increases, does the episodic component (η) of the temporal structure of the free recall protocol decrease while the causal component (β) increases or remains constant? And third, as age increases, does the episodic component (η) of the temporal structure of the free recall protocol increase while the causal component (β) decreases or remains constant?

Note that the above general qualitative pattern of results have been addressed previously by other researchers in the field. For example, researchers have shown that as retention interval increases, people tend to forget items with fewer causal connections [25,39,40]. In addition, researchers have shown that adults recall of a story is less influenced by the causal structure of the story than children's recall [30,31]. On the other hand, these previous studies did not show explicitly how the relative influences of semantic and episodic knowledge sources influence recall performance as a function of treatment condition. The proposed connectionist model directly addresses this question. Moreover, if these basic findings derived from conventional data analyses can be replicated using a categorical time-series analysis model, then this would support the hypothesis that information regarding the *order in which propositions are recalled from memory* is a potentially useful source of information that can be successfully exploited by the proposed analysis.

4.2.2. *Parameter Estimates*

Before beginning, it is necessary to explain the organization of the parameter vector ρ. Let ρ be constructed as in (4.1) such that the first group of the elements of the ρ vector correspond to parameters estimated from the children recall data (i.e., group one), the second group of the elements of the ρ vector correspond to parameters estimated from the immediate adults recall data (i.e., group two), and the third group of the elements of the ρ vector correspond to parameters estimated from the adults who recalled the texts after a delay of one week (i.e., group three). Each

group of elements of the ρ vector is a subvector of dimension 16 since each treatment group reads 4 different texts and 4 parameters are estimated for each text.

For reference, the parameter estimates for the twelve experimental conditions of the study are displayed in Table 3. Text 1 was the story *Epaminondas*, text 2 was the story *Judy's Birthday*, text 3 was the story *Tiger's Whisker*, and text 4 was the story *Fox and Bear*. Treatment condition 1 referred to the data from the fifth graders who immediately recalled the stories. Treatment condition 2 referred to the data from the adults who immediately recalled the stories. And treatment condition 3 referred to the data from the adults who recalled the stories one week after reading the stories. Additional details regarding the recall data and the digraph representations of the *Epaminondas* story may be found in the Appendix. The recall data for these four stories was obtained from [12].

Table 3

Condition	η_i^j	β_i^j	γ_i^j	λ_i^j
ρ_1^1	2.932	2.392	3.145	10
ρ_2^1	3.774	2.633	4.104	9
ρ_3^1	4.136	1.933	3.477	9.615
ρ_4^1	3.768	1.639	3.674	12.15
ρ_1^2	3.439	2.339	4.493	14.58
ρ_2^2	3.72	2.076	3.349	13.71
ρ_3^2	4.385	1.729	4.02	12.54
ρ_4^2	4.414	2.054	4.063	15.58
ρ_1^3	3.463	2.215	3.971	13.88
ρ_2^3	3.378	2.089	3.192	12.46
ρ_3^3	4.003	1.461	3.684	11.75
ρ_4^3	3.997	1.581	3.878	14.42

4.2.3. *Comparisons Between Adult Immediate and Delayed Recall Conditions*

The first of two planned comparisons was designed to see if a difference existed between the adult immediate recall condition and the adult delayed recall condition of the experiment across all four stories. Each of the two comparisons tested a different composite hypothesis. Using the Bonferroni inequality, the significance level for each comparison was set to be 0.025 in order to assure that the experiment-wise significance level was less than 0.05.

Let the row subvector $\mathbf{u}_1 = [1\ 0\ 0\ 0]$, the row subvector $\mathbf{u}_2 = [0\ 1\ 0\ 0]$, the row subvector $\mathbf{u}_3 = [0\ 0\ 1\ 0]$, and the row subvector $\mathbf{u}_4 = [0\ 0\ 0\ 1]$. Let the row subvector $\mathbf{0} = [0\ 0\ 0\ 0]$. Then, the selection matrix for this comparison is given by the formula:

$$\mathbf{S} = 0.25 * \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 & \mathbf{u}_4 & -\mathbf{u}_1 & -\mathbf{u}_2 & -\mathbf{u}_3 & -\mathbf{u}_4 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 & \mathbf{u}_4 & -\mathbf{u}_1 & -\mathbf{u}_2 & -\mathbf{u}_3 & -\mathbf{u}_4 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 & \mathbf{u}_4 & -\mathbf{u}_1 & -\mathbf{u}_2 & -\mathbf{u}_3 & -\mathbf{u}_4 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 & \mathbf{u}_4 & -\mathbf{u}_1 & -\mathbf{u}_2 & -\mathbf{u}_3 & -\mathbf{u}_4 \end{bmatrix}. \quad (4.4)$$

To understand the rationale for this choice of **S**, remember the null hypothesis of the Wald test is given by $H_o : \mathbf{S}\rho = 0$. Thus, the above choice of **S** yields the following null hypothesis:

$$H_o : \begin{cases} (\eta_1^2 + \eta_2^2 + \eta_3^2 + \eta_4^2)/4 = (\eta_1^3 + \eta_2^3 + \eta_3^3 + \eta_4^3)/4 \\ (\beta_1^2 + \beta_2^2 + \beta_3^2 + \beta_4^2)/4 = (\beta_1^3 + \beta_2^3 + \beta_3^3 + \beta_4^3)/4 \\ (\gamma_1^2 + \gamma_2^2 + \gamma_3^2 + \gamma_4^2)/4 = (\gamma_1^3 + \gamma_2^3 + \gamma_3^3 + \gamma_4^3)/4 \\ (\lambda_1^2 + \lambda_2^2 + \lambda_3^2 + \lambda_4^2)/4 = (\lambda_1^3 + \lambda_2^3 + \lambda_3^3 + \lambda_4^3)/4 \end{cases} \tag{4.5}$$

by (4.4), and using the parameters estimated for the jth text in the ith experimental condition: $\rho_j^i = [\eta_j^i \;\; \beta_j^i \;\; \gamma_j^i \;\; \lambda_j^i]$, and (4.1).

Let K_α be defined such that the probability that a chi-squared random variable with four degrees of freedom is greater than K_α with probability α. Using (4.3), a Wald statistic can be computed that has a chi-squared distribution with four degrees of freedom since the rank of the selection matrix **S** is of rank four [44]. If the Wald statistic, W, is greater than K_α, then the null hypothesis in (4.5) is rejected at the α significance level. The value of the Wald statistic in this case is 11.86 and has a significance level less than 0.025. Thus, the difference between the immediate and delayed adult recall conditions is significantly different. The results of the statistical analysis are summarized using the notation: $W(4) = 11.86, p < 0.025$.

To investigate further the nature of this between-groups difference, a post-hoc contrast was constructed to see if the average of the estimated causal strength parameters across the four stories recalled by the subjects in the immediate recall condition ($\beta = 2.05$) was significantly different from the average of the estimated causal strength parameters across the four stories recalled by the subjects in the 1-week delay recall condition ($\beta = 1.84$). A statistical test designed to examine this hypothesis can be constructed by choosing the selection matrix **S** to have the form:

$$\mathbf{S} = \mathbf{0.25} * [\mathbf{0} \;\; \mathbf{0} \;\; \mathbf{0} \;\; \mathbf{0} \;\; \mathbf{u_2} \;\; \mathbf{u_2} \;\; \mathbf{u_2} \;\; \mathbf{u_2} \;\; -\mathbf{u_2} \;\; -\mathbf{u_2} \;\; -\mathbf{u_2} \;\; -\mathbf{u_2}] \,. \tag{4.6}$$

Thus, when this selection matrix **S** is substituted into the expression: $H_o : \mathbf{S}\rho = \mathbf{0}$ this yields the null hypothesis:

$$H_o : (\beta_1^2 + \beta_2^2 + \beta_3^2 + \beta_4^2)/4 = (\beta_1^3 + \beta_2^3 + \beta_3^3 + \beta_4^3)/4 \,, \tag{4.7}$$

where β_j^i is the causal strength parameter value for the jth text in the ith experimental condition. Again, substituting the selection matrix **S** into (4.2) and (4.3), a statistical test can be constructed. In this case, the difference is not significant ($W(1) = 2.295, p = 0.1298$). On the other hand, the average of the estimated episodic strength (η) parameters across the four stories in the immediate recall condition ($\eta = 3.99$) was significantly greater than the average of the estimated episodic strength parameters across the four stories in the 1-week delayed recall condition ($\eta = 3.71$). It should also be noted that the average of the estimated causal strength parameters across the adults and children data was significantly

greater than zero ($\beta = 2.01, W(1) = 847.1, p < 0.0001$) extending and replicating previous findings [18,20,40,39] that causal structure does influence memory for text using the temporal structure of recall protocols as the dependent measure. Finally, the average number of text features recalled in the immediate adult recall condition ($\lambda = 14.1$) was greater than the average number of items recalled in the delayed recall condition ($\lambda = 13.13, W(1) = 5.725, p < 0.02$). The results reported in this section are consistent with previous findings that subjects tend to forget statements with fewer causal connections [39,40] as the time period between reading the story and recalling the story increased. These results also extend previous findings in two distinct ways. First, they demonstrate that the *order* of the propositions in a text provide a useful but frequently neglected source of information regarding the structure of underlying cognitive representations and that the proposed analysis is sensitive enough to exploit this novel source of information. And second, the model can statistically factor out the effects of causal text structure from effects of episodic text structure.

4.2.4. *Comparisons Between Children and Adults Data*

The second planned comparison was designed to test the null hypothesis that the parameter values estimated for the children averaged across stories were identical to the parameter values estimated for the adults immediate recall data averaged across stories. The selection matrix for this planned comparison was constructed in a manner similar to the selection matrix in (4.4). The results of the analysis demonstrated a significant difference between these two groups of subjects ($W(4) = 73.18, p < 0.0001$). Post-hoc analyses of the data indicated that these differences were due to two factors. First, the estimated average number of items recalled from each story averaged across stories for the children recall condition ($\lambda = 10.19$) was less than the estimated average number of items recalled from each story averaged across stories for the adult recall condition ($\lambda = 14.1$). This difference was significant ($W(1) = 67.76, p < 0.0001$). Second, the estimated episodic strength parameter averaged across stories for the children recall data ($\eta = 3.65$) was less than the estimated episodic strength parameter averaged across stories for the adult immediate recall data ($\eta = 3.99, W(1) = 4.87, p < 0.05$). The estimated causal strength parameter averaged across stories for the children recall condition ($\beta = 2.15$) was not significantly different from the estimated causal strength parameter averaged across stories for the adult immediate recall condition ($\beta = 2.05, W(1) = 0.3081, p = 0.58$).

These findings replicate the findings of Mandler [30] and Mandler and DeForest [31] that have indicated children are more likely to recall stories in their canonical order as opposed to the original order of presentation relative to adults. These findings also extend this previous work in three important ways. First, they suggest that the *order of the propositions in the recall protocol* are an important source of

information about cognitive representations. Second, they suggest that the model is sensitive enough to exploit this source of information. And third, these findings suggest children's recall is more schema based than adults's recall simply because children forget more details about the text's original episodic structure. Additional studies are currently planned to replicate these findings using more texts and better digraph representations.

5. Summary and Conclusions

A new method for the analysis of categorical time-series text recall data has been proposed. The method differs from conventional analyses in three important ways. First, each trajectory (i.e., recall protocol) is considered to be a single observation or data point. Second, the method provides for the incorporation of a considerable amount of prior knowledge regarding how the working memory contents of the model will predict the likelihood of the next feature in a trajectory. The method can also represent prior knowledge regarding alternative coherence strategies for deciding how long a feature should remain active in working memory. Third, statistical analyses for comparing alternative choices of prior knowledge structures and analyzing the effects of different treatment conditions on the parameter estimates are developed within a model misspecification framework [44,43], and therefore do not require the classical goodness-of-fit assumption to guarantee the reliability of the large sample statistical tests. The proposed model was then applied to the analysis of free recall data from four stories in order to explore the validity of the asymptotic approximations involving real data sets. The results of these analyses were encouraging and indicated the model could use the order of the recalled propositions from a text to: (i) discriminate among alternative causal knowledge structures and (ii) examine the nature of between-treatment differences in terms of interpretable parameter estimates.

References

[1] Allison P. D. and Liker J. K., Analyzing sequential categorical data on dyadic interaction: A comment on Gottman, *Psychological Bulletin* **91** (1982) pp. 393–403.

[2] Bartholomew D. J., *Latent Variable Models and Factor Analysis* (Oxford Univ. Press, New York, 1987).

[3] Black J. B. and Bern H., Causal coherence and memory for events in narratives, *J. Verbal Learning and Verbal Behavior* **20** (1981) pp. 267–275.

[4] Efron B., *The Jackknife, the Bootstrap and Other Resampling Plans.* Society for Industrial and Applied Mathematics (SIAM, 1982), Philadelphia.

[5] Efron B. and Tibshirani, *An Introduction to the Bootstrap* (Chapman and Hall, New York, 1993).

[6] Fletcher C. R. and Bloom C. P., Causal reasoning in the comprehension of simple narrative texts, *J. Memory and Language* **27** (1989) pp. 235–244.

[7] Gillund G. and Shiffrin R. M., A retrieval model for both recognition and recall, *Psychol. Rev.* **19** (1984) pp. 1–65.

[8] Golden R. M., Probabilistic characterization of neural model computations. In *Neural Networks and Information Processing*, ed. by Anderson D. Z. (AIP, New York, 1988) pp. 310–316

[9] Golden R. M., Relating neural networks to traditional engineering approaches. In *Proc. of the Artificial Intelligence and Advanced Computer Technology Conf.* (Tower Conference Management Company, IL, 1988) pp. 255–260.

[10] Golden R. M., A unified framework for connectionist systems. *Biological Cybernetics* **59** (1988) pp. 109–120.

[11] Golden R. M., Making correct statistical inferences using a wrong probability model, *J. Math. Psychol.* (in press).

[12] Golden R. M., Golden S. F., Strickland J. and Choi I., A psychometric pdp model of temporal structure in story recall. In *Proc. of the Fourteenth Ann. Conf. of the Cognitive Science Society* (Erlbaum, Hillsdale, NJ, 1993) pp. 487–491.

[13] Golden R. M. and Rumelhart D. E., A distributed representation and model for story comprehension and recall. In *Proc. of the Thirteenth Ann. Conf. of the Cognitive Science Society* (Erlbaum, Hillsdale, NJ, 1991) pp. 7–12.

[14] Golden R. M. and Rumelhart D. E., A parallel distributed processing model of story comprehension and recall. *Discourse Processes* **16** (1993) pp. 203–237.

[15] Golden R. M., Rumelhart D. E., Strickland J. and Ting A., Markov random fields for text comprehension. In *Neural Networks for Knowledge Representation and Inference* (Erlbaum, Hillsdale, NJ, 1994).

[16] Goodman L. A., *Analyzing Qualitative/Categorical Data* (Abt Books, Cambridge, MA, 1978).

[17] Gottman J. and Roy A. K., *Sequential Analysis: A Guide for Behavioral Researchers* (Cambridge Univ. Press, New York, 1990).

[18] Graesser A. C., How to catch a fish: The memory and representation of common procedures, *Discourse Processes* **1** (1978) pp. 72–89.

[19] Graesser A. C. and Clark L. F., *Structures and Procedures of Implicit Knowledge*, (Ablex, Norwood, NJ, 1985).

[20] Graesser A. C., Robertson S. P., Lovelace E. R. and Swinehart D. M., Answers to why questions expose the organization of story plot and predict recall of actions, *J. Verbal Learning and Verbal Behavior* **19** (1980) pp. 110–119.

[21] Haberlandt K. and Bingham G., The effect of input direction on the processing of script statements, *J. Verbal Learning and Verbal Behavior* **23** (1984) pp. 162–177.

[22] Hays W. L., *Statistics* (Holt, Rinehart and Winston, New York, 1963).

[23] SAS Institute, *SAS User's Guide: Statistics, Version 5 Ed.* (SAS Inst. Inc., Cary, NC, 1985).

[24] Kintsch W., The role of knowledge in discourse comprehension: a construction-integration model, *Psychol. Rev.* **95** (1988) pp. 163–182.

[25] Kintsch W. and VanDijk T. A., Toward a model of text comprehension and production, *Psychol. Rev.* **85** (1978) pp. 363–394.

[26] Kullback S. and Leibler R. A., On information and sufficiency, *Annals of Math. Stat.* **22** (1951) pp. 79–86.

[27] Levinson S. E., Ljolje A. and Miller L. G., Large vocabulary speech recognition using a hidden markov model for acoustic/phonetic classification. In *1988 IEEE Int. Conf. on Acoustics, Speech and Signal Processing, Vol. 1* (IEEE Service Center, Piscataway, NJ, 1988) pp. 505–508.

[28] Luce R. D., Detection and recognition. In *Handbook of Mathematical Psychology, Vol. 1*, ed. by Luce R. D., Bush R. R. and Galanter E., (Wiley, New York, 1963) pp. 103–189.

[29] Mackie J. L., *The Cement of the Universe: A Study of Causation* (Clarendon Press, Oxford, 1980).
[30] Mandler J. M., A code in the node: The use of a story schema in retrieval, *Discourse Processes* **1** (1978) pp. 14–35.
[31] Mandler J. M. and DeForest M., Is there more than one way to recall a story? *Child Development* **50** (1979) pp. 886–889.
[32] McKoon G. and Ratcliff R., Semantic associations and elaborative inference, *J. Experimental Psychology: Learning, Memory and Cognition* **15** (1989) pp. 326–338.
[33] Mensink G. and Raaijmakers J. G. W., A model for interference and forgetting, *Psychol. Rev.* **95** (1988) pp. 434–455.
[34] Murdock B. B., Direction of recall in short-term memory, *J. Verbal Learning and Verbal Behavior* **1** (1962) pp. 119–124.
[35] Murdock B. B., Forward and backward associations in paired associates, *J. Experimental Psychology* **71** (1966) pp. 732–737.
[36] Myers J. L., Shinjo M. and Duffy S. A., Degree of causal relatedness and memory, *J. Memory and Language* **26** (1987) pp. 453–465.
[37] Raaijmakers J. and Shiffrin R., Search of associative memory, *Psychol. Rev.* **88** (1981) pp. 93–134.
[38] Singer M. and Ferreira F., Inferring consequences in story comprehension, *J. Verbal Learning and Verbal Behavior* **22** (1983) pp. 437–448.
[39] Trabasso T., Secco T. and VanDenBroek P., Causal cohesion and story coherence. In *Learning and Comprehension of Text*, ed. by Mandl H., Stein N. L. and Trabasso T. (Erlbaum, Hillsdale, NJ, 1984) pp. 83–111.
[40] Trabasso T. and VanDenBroek P., Causal thinking and the representation of narrative events, *J. Memory and Language* **24** (1985) pp. 612–630.
[41] VanDenBroek P., The causal inference maker: Towards a process model of inference generation in text comprehension. In *Comprehension Processes in Reading*, ed. by Balota D. A., d'Arcais G. B. and Rayner K. (Erlbaum, Hillsdale, NJ, 1990) pp. 423–445.
[42] VanDenBroek P. and Lorch R. F., Network representations of causal relations in memory for narrative texts: Evidence from primed recognition, *Discourse Processes*, in press.
[43] Vuong Q. H., Likelihood ratio tests for model selection and non-nested hypotheses, *Econometrica* **57** (1989) pp. 307–333.
[44] White H., Maximum likelihood estimation of misspecified models, *Econometrica* **50** (1982) pp. 1–25.
[45] White H., Learning in artificial neural networks: A statistical perspective, *Neural Computation* **1** (1989) pp. 425–464.
[46] White H., *Estimation, Inference and Specification Analysis* (Cambridge Univ. Press, New York), in press.
[47] Wilks S. S., The large sample distribution of the likelihood ratio for testing composite hypotheses, *Annals of Math. Stat.* **9** (1938) pp. 60–62.
[48] Stein N. L. and Glenn C. G., An analysis of story comprehension in elementary school children. In *New Directions in Discourse Processing*, ed. by Freedle R. O. (Norwood, NJ: Ablex, 1979).

Appendix: Digraph Representation of *Epaminondas* Text

The purpose of this appendix is to provide additional details regarding the digraph models used in the analyses of the *Epaminondas* story (text 1) discussed in the article. Additional details regarding the other three texts which were analyzed can be obtained by contacting the author.

Table 4. Text feature table [48].

Feature Table (Text 1)	
Feature ID	Sentence Representation
f_1	The boy was little.
f_2	The boy lived in a hot country.
f_3	The mother told the boy to take the cake to his grandmother.
f_4	The mother told the boy to hold the cake carefully.
f_5	The mother told the boy to prevent the cake from breaking into crumbs.
f_6	The boy put the cake in a leaf under his arm.
f_7	The boy carried the cake to his grandmother.
f_8	The boy arrived at his grandmother's house.
f_9	The cake had crumbled.
f_{10}	The grandmother told the boy he was silly.
f_{11}	The grandmother told the boy he should have carried the cake on his head.
f_{12}	The grandmother told the boy he could have prevented the cake from crumbling.
f_{13}	The grandmother told the boy to take butter to his mother.
f_{14}	The boy held the butter carefully.
f_{15}	The boy put the butter on his head.
f_{16}	The boy carried the butter home.
f_{17}	The sun was shining hard.
f_{18}	The boy arrived home from his grandmother's house.
f_{19}	The butter was melted.
f_{20}	The mother told the boy he was silly.
f_{21}	The mother told the boy to put the butter into a leaf.
f_{22}	The mother told the boy to keep the butter safe.

Table 5. Episodic digraph model.

Episodic Digraph (Text 1)	
To Feature	From Feature
2	1
3	2
4	3
9	8
12	11
13	12
15	14
16	15
17	16
18	17
19	18
21	20

Table 6. Causal digraph model.

Causal Digraph (Text 1)	
To Feature	From Feature
6	3
6	4
7	3
8	3
9	6
9	7
10	6
11	9
12	9
15	13
16	13
18	13
19	15
19	16
19	17
20	15
21	16
21	17
21	19

Table 7. Shared causal-episodic digraph model.

Shared Digraph (Text 1)	
To Feature	From Feature
5	4
6	5
7	6
8	7
10	9
11	10
14	13
20	19
22	21