Multivariate Behavioral Research, *38* (4), 463-504

Multivariate Multinomial Logit Models for Dyadic Sequential Interaction Data

Mark de Rooij
Leiden University

Pieter M. Kroonenberg
Department of Education
Leiden University

The analysis of discrete dyadic sequential behavior and, in particular, the problem of forecasting future behavior from current and past behavior in such data is the main theme of the present article. We propose to use multivariate multinomial logit models and the potential of which will be demonstrated with data on Imagery play therapy. In such a therapy, the therapist tries to draw a child into Imagery play, so that it can act out its emotions and feelings which gives the therapist the opportunity to communicate with the child. As the therapist wants to interact clinically with the child, it is important to draw it into Imagery play as soon as possible. Our aim is to find out how the therapist achieves this by examining and forecasting new behavior from past behavior. New behavior can be modeled by special weighting schemes, but this procedure can be technically problematic. In this paper the nature of these problems will be explained and methods to solve them will be proposed. Moreover, we will explicitly attempt to answer the substantive questions for which the data were collected through a detailed interpretation of the parameters of the models.

Introduction

In the behavioral sciences one regularly encounters studies in which the interaction between two subjects (A and B) is observed over a period of time. Examples are studies by Harinck and Hellendoorn (1987), who observed a therapist and a child in Imagery play therapy; by Van de Boom (1988), who observed the interaction between mother and child; by Benes, Gutkin, and Kramer (1995), who studied communication processes in school-based

The Netherlands Organization for Scientific Research (NWO) is gratefully acknowledged for funding this project. This research was conducted while the first author was supported by a grant of the Foundation for Behavioural and Educational Sciences of this organization (575-30-006). The authors are indebted to Frits Harinck and Joop Hellendoorn for providing their data and useful comments on the interpretation of obtained results. Requests for reprints should be sent to Mark de Rooij, Department of Psychology, Leiden University, PO Box 9555, 2300 RB Leiden, THE NETHERLANDS. E-mail: rooijm@fsw.leidenuniv.nl

consultation; by Voeten (1985), who observed a class room, and paid attention to the interaction between the teacher and his or her pupils, and the studies of interaction of married couples by, among others, Gottman (1979a, 1979b).

Models for Dyadic Sequential Data

The last two decades have also seen a growing attention to the development of statistical models for such dyadic sequential data (Allison & Liker, 1982; Bakeman & Gottman, 1997; Bakeman & Quera, 1995; Budescu, 1984; Gottman, 1979a, 1979b, 1981; Iacobucci & Wasserman, 1987, 1988). The trend in this series of articles is towards a log-linear approach for the analysis of sequential dyadic data. However, as Bakeman and Quera noted: "But only rarely have researchers with sequential data used log-linear analysis even though such analyses seem well suited for such data" (p. 273).

The reason for not using standard log-linear models is probably the difficulty of interpreting the results of such an analysis. In practice, we see that the parameters of the log-linear model are not even estimated, but only the expected frequencies, and interpretation of the results is carried out by looking at (standardized) residuals from particular models rather than interpreting the parameters of these models.

The parameters of a log-linear model are difficult to interpret for a number of reasons. First of all, there is no underlying empirical scale of the dependent variable. Second, the number of parameters that must be interpreted is very large. Because of these difficulties, often the parameters of a model are not even estimated but only the important association effects are determined with the corresponding expected frequencies, so little can be said about the direction of association. In this article the focus will be on the first and third problem. The second problem of interpreting many parameters can be handled by taking a graphical approach using correspondence analysis (e.g., Greenacre, 1984) or techniques based on distances (see de Rooij, 2001).

There are basically two alternatives (close to the log-linear model) to deal with the problem of interpretability. Budescu (1984) adopted the weighted least squares approach proposed by Grizzle, Starmer, and Koch (1969) using an interpretation strategy similar to standard regression analysis and analysis of variance. A second approach, which will be discussed in the present article, is a relatively straightforward reparametrization of the log-linear model (see Haberman, 1979) that leads to an interpretation in terms of odds and odds ratios. The model is often called multinomial response model or multinomial logit model. The first and most often encountered description of the multinomial logit model can be found in Theil (1969, 1970). Recent extensions to ordered categorical dependent variables like the cumulative logit models or proportional odds models can be found in an influential paper by McCullagh (1980). Multinomial logit models can be interpreted by taking a latent variable perspective (see Long, 1997, chapter 6; Powers & Xie, 2000, chapter 7). A conditional approach to the multinomial logit model is described in the Conditional Logit Model, also known as Luce model or Discrete choice model (McFadden, 1974). In the Conditional Logit Model the coefficients for a variable are the same for each outcome, but the values of the variables differ for each outcome.

We will briefly show how the reparametrization from log-linear towards multinomial logit models works, then deal extensively with the interpretation of parameter estimates, and show how we can relate results of the analysis to concepts as direct and indirect dominance (Allison & Liker, 1982; Budescu, 1984; Gottman, 1979b). This reparametrization makes a distinction between explanatory and response variables, which gives the model regression features and allows for a forecasting interpretation in terms of (ratios of) conditional probabilities. In the section entitled "Results" we will show an example in which this forecasting mode is important.

Sampling Schemes for Dyadic Sequential Data

For the analysis of two-way square contingency tables that are indexed by time an important class of models is the mover-stayer models (Hout, 1983). In these models the diagonal cells and the off-diagonal cells are treated separately, where the diagonal cells correspond to the stayers, the off-diagonal to the movers. Goodman (1968) showed that the mover-stayer model can be estimated by designating the diagonal cells as having structurally missing observations, generally referred to as *structural zeroes*. The assumption is that no observations can be recorded in such cells due to the nature of the design. By using this device, the analyses only apply to the off-diagonal elements and are thus specific for the movers. In dyadic sequential data we can also distinguish between movers and stayers, although the design is somewhat more complicated. Movers can be specified in terms of the dyad or in terms of one of the actors. This relates to different sampling schemes. First, we deal with the movers and stayers (*time-sampling scheme*), then with movers defined on the dyad (*event-sampling scheme*), and finally with movers based on one actor (*actor-specific event sampling scheme*). For an extensive discussion of observational schemes see Bakeman and Gottman (1997, chapter 3).

Time Sampling

An often used observational scheme in sequential research is time sampling, that is, the two actors are observed and each *t* seconds their behavior is recorded. When using such a scheme, it is important that the interval between two consecutive observations is small enough such that all relevant behavior is captured. An example of a dyadic series of observations is given in the upper part of Table 1, where two actors are observed and their behavior is categorized into four categories, 1, 2, 3, and 4. The specific behavioral category 1 of actor *A* is denoted as a_1, and at a specific time point, for example t_{-1}, as a_1^{-1}. From such a sequence of observations a contingency table can be formed by using a moving time window. Therefore, the first transition from time point 1 to time point 2 ($a_1^{-1}b_1^{-1} \rightarrow a_1^0 b_2^0$) is observed and tallied in the corresponding cell of the contingency table relating the behavior of both actors at t_{-1} and t_0. Then the window is moved one place, and the next transition ($a_1^{-1}b_2^{-1} \rightarrow a_1^0 b_2^0$) is tallied. This process continues until the last transition is tallied.

Table 1
An Example of Dyadic Sequential Data with Different Sampling Schemes

									Time sampling											
A	1	1	1	1	1	1	4	4	4	4	2	2	3	3	3	1	1	1	1	…
B	1	2	2	2	4	4	4	4	1	1	1	2	2	3	3	3	1	4	4	…
									Event sampling											
A	1	1			1		4		4		2	2	3	3		1	1	1		…
B	1	2			4		4		1		1	2	2	3		3	1	4		…
									Actor-specific event sampling											
A						1	4			4	2	2	3		3	1			…	
B						4	4			1	1	2	2		3	3			…	

Event Sampling

When the interest is mainly on changes of behavior, a disadvantage of the time-sampling scheme and the way of tallying is that the entries in the contingency tables corresponding to auto-transitions (for example, $a_1^{-1}b_2^{-1} \rightarrow a_1^0 b_2^0$) are relatively large. Another observational scheme is event sampling, where behavior is only recorded when it changes. Entries of the contingency table related to auto-transitions are then structurally zero. However, given a time-sampling scheme was used to collect the data, and the primary interest is in changes of behavior, weights can be used during modeling to ignore cells which do not correspond to changes in behavior. By attaching zero weights to these cells, the data acquire the same structure as in event sampling. In which case, the first series of Table 1 reduces to the second series.

Actor-Specific Event Sampling

In some research designs, one of the two partners is of specific interest. An example is given in detail later, where a therapist and child are observed during a play-therapy session, and where the main interest is in the behavior and changes of behavior of the child. Such a design could be implemented by introducing a special observational scheme, that is, record the behavior of both actors when this one actor (the child) makes a change. For example, on the first two lines in Table 1 the first change of actor *A* is a change from category 1 to category 4, that will be the first observation together with the behavior of actor *B* that is tallied. This scheme will be called actor-specific event sampling. This can also be obtained by applying a weighting scheme, where all cells that correspond to auto-transitions of one actor are given a zero weight. The sequence of the first two lines of Table 1 then reduces to the sequence shown in the last two lines in the case of actor-specific event sampling, where only behavior is tallied that is within two vertical lines.

Focus of this Article

Kroonenberg and Verbeek (1990) showed that dealing with structural zeros or zero weights in multinomial logit models is extremely problematic in most popular statistical packages. One of the major aims of this article is to show how the estimation can be done in a correct and natural way.

Summarizing the main methodological focus of the present article:

1. We propose to use multinomial logit models for the analysis of dyadic sequential data as they have interpretational advantages over the standard log-

linear approach. This advantage will be used to answer substantive questions related to the data, and the interpretations will be carried out via a detailed discussion of the parameters of selected models for each of the three sampling schemes.

2. We will show that the computational problems in the case of structural zeros when estimating the parameters of multinomial logit models can be solved and we will show how valid conclusions can be obtained.

Method

Data and Subjects

The substantive problems guiding our methodological considerations and modeling originated with Harinck and Hellendoorn (1987) in the context of Imagery play therapy. In this kind of therapy a child and a therapist play for approximately an hour. The central ingredients of the therapeutic interaction are the play activities. The therapist tries to draw the child into so-called Imagery play, because this kind of play allows for the expression of personal experiences (Kroonenberg, Hellendoorn, & Harinck, 1988), and thus gives the therapist a handle to discuss its problems. By using Imagery play, the child can express its problems with its parents in a make-believe world, without having the idea of being disloyal to its parents.

The observational scheme for this kind of therapy is as follows. A child and therapist are observed during a therapeutic session of approximately an hour, and their behaviors are scored every five seconds (time sampling). The original scores were coded using 34 categories for both therapist and child. We recoded, following Kroonenberg, Hellendoorn, and Harinck (1988), the 34 categories into four major categories, Non play (N), Play preparation (P), Functional play (F), and Imagery play (I). For the child the category Non play consists of activities as talking, giving support, giving attention, and activities not covered by the other categories. For the therapist the Non play category consists for 75% of so-called 'simple attention', that is, the therapist watches the activities of the child. The purpose of this activity is to give the child the opportunity to develop its play without the therapist's inference, and according to Hellendoorn (personal communication) this is an important element in the therapy. Play preparation consists of activities as selection and naming of material, discussion of its play possibilities, discussion of play subjects, and arranging things without special meaning. Functional play covers movement with play things in the way their properties dictate (roll a ball, ride a car), and also emphasizes special sensory and esthetic qualities of play. Finally, Imagery play consists of activities as creating a play world or a play situation from which play things derive their special meaning, playing a story or event, adding a special feature to the imaginative situation or event, emphasizing the feelings or emotions of play persons, evaluative comments and interpretive remarks in play or giving therapeutic messages in play from.

In the sequence of play scores we often see Non play and Play preparation as initial stages of play. From these initial stages, play develops to either Functional play or Imagery play. Not many transitions occur between the two latter kinds of play. They are so-called 'end-points' in a play sequence. Once in Functional play, a child generally goes back to one of the initial stages (N or P), before a transition towards Imagery play can be made (Harinck & Hellendoorn, 1987). This makes it infeasible to deal with the categories as an ordered set. The main question in Harinck and Hellendoorn's research was which combination of play of the child and the therapist will lead to future Imagery play of the child. In other words, which categories of child and therapist at t_{-1} (and earlier times) lead to the highest probability of Imagery play of the child at t_0.

A total of 117 dyads of therapists and children (48 children in therapy and 69 children not in therapy) were observed, yielding over 70,000 dyadic scores. For each dyad a contingency table is made by tallying consecutive behavior, afterwards all contingency tables are added to obtain one overall transition frequency table, Table 2. In doing so, we implicitly assume two things: First, we assume homogeneity across all dyads; Second, we assume the behavior at t_0 can be predicted from the behavior at previous time points and that this predictability is stable over the whole sequence. We realize these assumptions are a severe simplification of the process, but we think it is in line with the research question, because we would like to find out how to draw future children (about whom we know nothing in advance) to the Imagery play mode. In principle, it is possible that the aggregated results do not reflect the pattern of any specific child, that is by pooling all dyads there exists a danger for Simpson's paradox. From a modeling point of view, in many cases like this one, it is practically impossible to model individual differences because of the sparseness of the data. Moreover, it was not our intention to model differences between children, but we will discuss how this might be done in the final section.

Unfortunately there were only four therapists. This makes that their therapy styles may have influenced the results, and that the data are non-independent. However, the therapists were equally distributed over the children. For the non-therapy children the distribution over the therapists is 18, 16, 18, and 17 (i.e., the first therapist played with 18 children, the second therapist with 16 children, etc.). Each of the therapists saw 12 therapy children. Generally, there exists non-independence in dyadic sequential data, since the observations in one pair for the

Table 2
Dyadic Sequential Interaction Data of Hellendoorn and Harinck

		c_1 d_1	d_2	d_3	d_4	c_2 d_1	d_2	d_3	d_4	c_3 d_1	d_2	d_3	d_4	c_4 d_1	d_2	d_3	d_4
a_1	b_1	3254	1019	338	154	292	407	53	14	84	48	83	4	128	91	39	94
	b_2	877	5867	591	475	406	2726	201	61	56	291	146	7	152	874	93	262
	b_3	332	532	1437	111	68	147	239	9	70	108	520	9	38	107	227	69
	b_4	149	357	89	2709	34	92	11	93	6	8	12	48	227	357	112	2671
a_2	b_1	322	348	58	21	526	687	74	34	51	41	39	2	49	69	6	38
	b_2	361	2582	212	148	643	4374	214	90	39	211	105	5	84	483	57	100
	b_3	47	147	266	13	77	162	291	12	24	55	184	2	18	25	57	14
	b_4	8	48	5	107	23	64	5	131	1	1	3	10	15	46	11	146
a_3	b_1	87	58	57	4	39	60	29	1	80	52	89	2	20	7	12	11
	b_2	55	329	98	10	30	212	40	3	38	234	147	4	13	69	15	14
	b_3	98	160	423	19	35	118	153	4	106	168	1095	14	19	46	127	28
	b_4	4	6	5	60	3	2	4	3	3	2	6	43	15	9	8	97
a_4	b_1	179	182	31	224	50	81	11	22	19	9	20	7	445	201	78	605
	b_2	149	1046	99	298	73	437	19	25	10	55	28	10	209	1304	167	484
	b_3	56	116	239	108	18	47	39	5	10	21	111	3	104	156	658	224
	b_4	160	270	71	2516	39	87	10	117	11	6	24	100	627	569	249	8967

Note. a = behavior of therapist at t_{-1}; b = behavior of child at t_{-1}; c = behavior of therapist at t_0; and d = behavior of child at t_0. Underlined frequencies are replaced by structural zeros in case of actor-specific event sampling.

beginning of a chain and the end of a chain are dependent in the same way. The chi-square statistics can be computed, but we should not rely heavily on the p-values and the theoretical distribution of the statistic. Still we can use such statistics as a descriptive device.

Multivariate Regression Models for Contingency Tables

In this section we will discuss a reparametrization of the log-linear model by distinguishing between explanatory (previous behavior) and response variables (current behavior), which allows for an interpretation of the parameters in terms of (log-)odds. We have used multinomial logit models for the analysis of dyadic sequential data, as was proposed earlier by, for instance, Gottman and Roy (1990, chapter 13). A detailed discussion of multinomial response models and multinomial logit models can be found in Agresti (1990, chapter 9), or Haberman (1979, chapter 6). The therapy data, however, consist of two response variables, thus in order to model these, we needed a multivariate multinomial logit model. Even though the idea of such models is not new, their use seems to be highly restricted as we could not find any appropriate references dealing with truly multivariate multinomial logit models, that is, models which had both more than one dependent variable which had at the same time more than two categories. An additional complication with the present data is that they required modelling in the presence of structurally zeroes which is far from straightforward in multivariate multinomial logit models, and poses several computational problems. Our discussion of multinomial logit models is based on the log-linear model, in contrast with the more usual approach where the multinomial logit model is seen as a generalization of the binary logit model (see, for instance, Long, 1997, chapter 6; Powers & Xie, 2000, chapter 6). To describe the models of interest the following notation is used: The categories of behavior for actor A will be denoted by a_k^0 for Lag 0, and a_i^{-1} for Lag 1 (i, $k = 1, ..., K$). The categories of behavior for actor B will be denoted by b_l^0 for Lag 0 and b_j^{-1} for Lag 1 (j, $l = 1, ..., L$).

Multivariate Multinomial Response Model

A multivariate model for the log of the probability of behavior of both actors at Lag 0 conditional on their behavior at Lag 1 (alternatively this could be Lag T, or Lag 1 and Lag 2, etc.) is needed. Such models are linear in the predictors. With all possible effects included (i.e., the saturated case), the model has the following form

$$
\begin{aligned}
\log \pi_{a_k^0 b_l^0 | a_i^{-1} b_j^{-1}} &= \gamma_{a_i^{-1} b_j^{-1}} + \lambda_{a_k^0} + \lambda_{b_l^0} + \lambda_{a_k^0 b_l^0} \\
&+ \lambda_{a_k^0 a_i^{-1}} + \lambda_{b_l^0 b_j^{-1}} && \text{auto-dependence} \\
&+ \lambda_{a_k^0 b_j^{-1}} + \lambda_{b_l^0 a_i^{-1}} && \text{cross-dependence} \\
&+ \lambda_{a_k^0 a_i^{-1} b_j^{-1}} + \lambda_{b_l^0 a_i^{-1} b_j^{-1}} && \text{interaction-dependence} \\
&+ \lambda_{a_k^0 b_l^0 a_i^{-1}} + \lambda_{a_k^0 b_l^0 b_j^{-1}} && \text{actor-dependence} \\
&+ \lambda_{a_k^0 b_l^0 a_i^{-1} b_j^{-1}} && \text{full-dependence}
\end{aligned} \tag{1}
$$

This is basically the same model as a log-linear model where the marginals of the explanatory variables are fixed by the γ-parameters to fit their observed values (cf. Fienberg, 1980, p. 95ff.). We will refer to this model the *multivariate multinomial response model*. If we set $\gamma_{a_i^{-1} b_j^{-1}} + \lambda_{a_i^{-1}} + \lambda_{b_j^{-1}} + \lambda_{a_i^{-1} b_j^{-1}}$, we get the standard log-linear formulation. The following groups of parameters can be distinguished: On the first line (together with the γ-parameter) parameters for the response variables are given, the next line gives parameters for *auto-dependence effects*, that is, an effect of previous behavior of actor A on itself and the same for actor B; the third line gives *cross-dependence effects*, the influence of the behavior of actor B on the behavior of actor A, and vice versa. The fourth group of parameters are labeled *interaction-dependence*, and denote a joint effect of both actors on one actor. *Actor-dependence* denotes the effects of one actor on the future interaction of the dyad. The last line is the four-way effect, that is, the effect of the interaction of the two actors on their future interaction.

Multivariate Multinomial Logit Model

The log-transformation accomplishes that differences of two log-probabilities can be understood as the log-odds given the behavior at the previous time point. This can be realized directly by using one behavioral category of both actors as baseline category (a_K^0, b_L^0), against which the others are contrasted. The multivariate multinomial response model described above can then can be written as a *multivariate multinomial logit model*:

$$
\begin{aligned}
\log\left(\frac{\pi_{a_k^0 b_l^0 | a_i^{-1} b_j^{-1}}}{\pi_{a_K^0 b_L^0 | a_i^{-1} b_j^{-1}}}\right) = \;& \alpha_{a_k^0} + \alpha_{b_l^0} + \alpha_{a_k^0 b_l^0} \\
+ \;& \beta_{a_k^0 a_i^{-1}} + \beta_{b_l^0 b_j^{-1}} && \text{auto-dependence} \\
+ \;& \beta_{a_k^0 b_j^{-1}} + \beta_{b_l^0 a_i^{-1}} && \text{cross-dependence} \\
+ \;& \beta_{a_k^0 a_i^{-1} b_j^{-1}} + \beta_{b_l^0 a_i^{-1} b_j^{-1}} && \text{interaction-dependence} \\
+ \;& \beta_{a_k^0 b_l^0 a_i^{-1}} + \beta_{a_k^0 b_l^0 b_j^{-1}} && \text{actor-dependence} \\
+ \;& \beta_{a_k^0 b_l^0 a_i^{-1} b_j^{-1}} && \text{full-dependence}
\end{aligned}
\tag{2}
$$

for $k = 1, ..., K - 1$, and $l = 1, ..., L - 1$, where $\alpha_{a_k^0} = \lambda_{a_k^0} - \lambda_{a_K^0}$, and $\beta_{a_k^0 a_i^{-1}} = \lambda_{a_k^0 a_i^{-1}} - \lambda_{a_K^0 a_i^{-1}}$, and similarly for the other αs and βs. The same groups of parameters can be distinguished as in the multinomial response model, but the interpretation of each parameter in model of Equation 2 is in terms of the odds of a category against the baseline category. Given the estimates of the model of Equation 2 we can compute any other log-odds since

$$
\log\left(\frac{\pi_{a_1^0|--}}{\pi_{a_2^0|--}}\right) = \log\left(\frac{\pi_{a_1^0|--}}{\pi_{a_K^0|--}}\right) - \log\left(\frac{\pi_{a_2^0|--}}{\pi_{a_K^0|--}}\right).
$$

For a detailed discussion of odds ratios the reader is referred to Rudas (1998).

Regression Models for Different Sampling Schemes

Time Sampling

With the multivariate multinomial logit model discussed in the previous section, a model was constructed for the contingency table with all observations resulting from the time sampling scheme. As time is the basic measuring unit, and the behaviors of the therapist and child showed considerable continuity, this continuity tended to dominate the modeling effort. However, the central research question was which behaviors of the therapist and child lead to Imagery play of the child, which question refers to change rather than continuity.

Event Sampling

As discussed in the introduction, an observational scheme may be employed in which only changes are observed (event sampling). In truly event sampling, the continuity cells of the contingency table are structural zeros, that is no observations are made for these cells. When time sampling has been employed and we are only interested in change, we want to treat continuity cells in the contingency table as structural zeros.

For event sampling schemes, the same models can be employed as in time sampling, but only those cells which refer to discontinuity, $\{\mathcal{S}\}$, are modeled. The continuity observations are treated as missing or are given a zero weight. In the latter case, the model is equal to the model of Equation 2, but only for a predefined set of cells $i, j, k, l \in \{\mathcal{S}\}$, where i and k relate to Lag 1 and Lag 0 behavior of actor A, respectively, and j and l to Lag 1 and Lag 0 behavior of actor B, respectively. In terms of weights, we define a set of weights w_{ijkl} such that $w_{ijkl} = 1$, if $i, j, k, l \in \{\mathcal{S}\}$, and $w_{ijkl} = 0$, if $i, j, k, l \notin \{\mathcal{S}\}$.

When the focus is on changes of the dyad (event sampling), the set $\mathcal{S}$ exists of all cells for which $(i \neq k) \wedge (j \neq l)$. The resulting four-way table for event sampling is equal to Table 2, where the cells on the main diagonal are given zero weight or replaced by structural zeros.

Actor-Specific Sampling

Another way to specify $\mathcal{S}$ is by focusing on changes of one actor (actor-specific event sampling) irrespective of the behavior of the other actor. In that case $\mathcal{S}$ consists of all cells for which $j \neq l$ is true. The resulting four-way table for actor-specific event sampling is equal to Table 2 where each underlined frequency is treated as a structural zero or have been given zero weights.

Model Selection, Evaluation, and Interpretation

Selection of Predictors

The interest is not in the full model as described in Equation 2 but in submodels where some effects are constrained to be equal to zero. In the case of a four-way table as discussed above, the number of possible models is already immense, and if one more lag is added a six-way table is obtained and the number of potential models increases significantly. However for dyadic sequential data, some guidelines exist which allow for a relatively systematic model selection procedure.

The first guideline is that before cross-dependence effects are fitted, auto-dependence effects should be included in the model. The reason is that cross-dependence effects might be induced by auto-dependence effects (Bakeman & Quera, 1995). To really say something about the cross-dependence effect one should therefore control for auto-dependencies in the data. Furthermore only hierarchical models are discussed, since higher-order effects can only be interpreted in addition to lower-order effects. Given a model, it is important to determine whether one of the partners is (in)directly dominant (see the section entitled "Differences Between Sets of Parameters") or not. Both test-statistics to deal with this question and parameter interpretation will be discussed.

Model Selection

In this section a number of fit statistics that are based on different idea's of fit are discussed. In general we believe that model selection should never be based on a single statistic. Model selection is a difficult process and therefore screening of multiple statistics can be helpful. We start with the traditional goodness-of-fit statistics, the likelihood ratio statistic

(3)
$$LR = 2\sum f_{ijkl} \log \frac{f_{ijkl}}{\hat{F}_{ijkl}},$$

where f_{ijkl} is the observed frequency for cell i, j, k, l and $\hat{F}_{ijkl}$ the estimated cell frequency for the same cell, and the Pearson's chi-square:

(4)
$$X^2 = \sum \frac{\left(f_{ijkl} - \hat{F}_{ijkl}\right)^2}{\hat{F}_{ijkl}}.$$

It is well known that both these statistics are a function of the sample size. When the number of observations is large these statistics tend to dismiss all models. Both statistics are not effect-size measures, but to obtain such measures X^2 or LR could be divided by the number of observations. An index which is not dependent on the sample size is the *dissimilarity index* (Agresti, 1996, p. 162):

(5)
$$D = \sum \frac{| f_{ijkl} - \hat{F}_{ijkl} |}{2N},$$

where N is the sample size. D can be interpreted as the proportion of observations which have to be moved to another cell in order for the specific model to fit the data perfectly, $1 - D$ is then the proportion correctly classified. In prediction or forecasting, this statistic is important in the sense that it gives an idea about the quality of our forecast. It can also be used as a measure of fit, in case a priori a maximum is specified for the proportion of the observations that has to be moved to obtain a perfect fit.

The likelihood ratio statistic, the Pearson chi-square statistic, and the dissimilarity index compare the observed frequencies with the expected frequencies. Using these statistics the saturated model is always the best model, since the expected frequencies in that case are equal to the observed frequencies. However, saturated models are not really interesting since such models have the same number of parameters as the number of cells, that is there is no reduction of information.

A different set of fit indices is obtained by making use of information theory. The *AIC*- and *BIC*-statistics are resultants from this approach. These statistics search for models that fit the data reasonably well with as few parameters as possible. Since both statistics behave in a similar way, but the *BIC*-statistic gives more weight to parsimony, the *BIC*-statistic is used. Raftery (1995) presented a nice and thorough discussion on the derivation and justification of the *BIC*-statistic. The *BIC*-statistic is defined as

$$BIC = LR - \log(N) \times df \tag{6}$$

where df is the degrees of freedom under the model fitted. The lower the *BIC*-statistic the better the model. In general, if a *BIC*-statistic smaller than zero is found, the model fits comparatively better than the saturated model, since the *BIC*-statistic is typically zero for the saturated model. A model with a negative value for the *BIC*-statistic then gives the same amount of information but using fewer parameters.

In regression analysis and analysis of variance the R^2-statistic or the percentage of variance explained is an important statistic. For categorical data, a number of pseudo R^2-measures have been proposed. Wickens (1989, pp. 127-132) gives a detailed discussion on the derivation of these measures. Since they all behave in a similar way, only the measure based on *concentration* is discussed here, which is defined for a one-dimensional probability distribution as

$$C(\pi) = 1 - \sum_j \pi_j^2.$$

If all probability falls into a single cell the concentration is zero, and the concentration is maximized when all categories are equally likely. For a table of observations under the product multinomial distribution, the concentration for a given model for row i is

$$C(\hat{\pi}_i) = 1 - \sum_j \hat{\pi}^2_{j|i},$$

and the concentration for the whole matrix is defined by a weighted average of the row concentrations, that is,

$$C(\hat{\pi}) = \sum_i \frac{f_{i+}}{f_{++}} C(\hat{\pi}_i),$$

where a + in the indices means summation over the index it replaces. An R^2-measure is obtained by comparing the concentration under a null model $C(\hat{\pi}_0)$ with the concentration of a model of interest $C(\hat{\pi}_{\mathcal{M}})$, in the following way

$$R^2_C(\mathcal{M}) = \frac{C(\hat{\pi}_0) - C(\hat{\pi}_{\mathcal{M}})}{C(\hat{\pi}_0)}. \quad (7)$$

Wickens (1989, p. 133) stated that models in general reduce the dispersion by only about 20%. So, this percentage is about the maximum that can be expected for our models.

Differences Between Sets of Parameters

Budescu (1984) defined *direct dominance* as: 'If under the appropriate model for predicting the joint behavior of a couple of actors from their past behavior, B's *future responses* are more predictable from A's past behavior than conversely, then A is said to be directly dominant.' In other words, to answer the question which actor is directly dominant the two sets of cross-dependence parameters from the model of Equation 2 have to be compared. *Indirect dominance* is defined by Budescu as: 'If under the appropriate model for predicting the joint behavior of a couple of actors from their past behavior, the nature of their *future interaction* is more predictable from A's behavior than from B's behavior, then A is said to be indirect dominant.' In other words, to answer the question which actor is indirect dominant the two sets of actor-dependence parameters from the model of Equation 2 have to be compared. To answer questions about direct dominance and indirect dominance (cf. Allison & Liker, 1982; Budescu, 1984; Gottman, 1979b), the strength of two sets of parameters need to be compared. Provided that a good fitting multivariate multinomial logit model is found, a test of the strength of a set of parameters $\gamma = (\gamma_1, ..., \gamma_q)'$, is given by the *Wald statistic* (Wald, 1943)[1]

$$W = \gamma'[\text{cov}(\gamma)]^{-1}\gamma, \tag{8}$$

which has a Chi-square distribution with $df = q$. In general, the difference between two Wald statistics can be tested as follows: Suppose W_1 and W_2 are two independent and identically distributed statistics, both chi-square distributed with q_1 and q_2 degrees of freedom respectively, then the ratio

$$\frac{W_1 / q_1}{W_2 / q_2}$$

has an F-distribution with q_1 and q_2 degrees of freedom (Mood, Graybill, & Boes, 1974, pp. 246-247). A problem in our application is that both χ^2-statistics come from the same model, and thus they are not independent. Nevertheless, the F-value can be used as an indication whether there exists dominance or not: if the numerical value of the ratio exceeds for example 6 (or is smaller that $^1/_6$) it is reasonable to assume that one is larger than the other. In our case this would be an indication of dominance.

Interpretation of Parameters

For both dependent variables a baseline category is defined in Equation 2. The effects can be understood as log-odds of a specific category against the baseline category. For example, the α-parameters denote the log-odds of an observation k against an observation K of actor A. Thus it is the probability of an observation k divided by a probability on an observation K. The β parameters have the same interpretation in terms of odds, but based on conditional probabilities. For example, $\beta_{a_k^0 a_i^{-1}}$ gives (on top of the α) the increase of the odds of an observation k against an observation K for actor

[1] The Wald statistic tests the null hypothesis H_0: $\gamma = \mathbf{0}$, against the alternative hypothesis this is not true. Like the Pearson X^2 and Likelihood ratio statistic, the Wald statistic is dependent on the sample size. Since we use the Wald statistic to compare (the strength of) two sets of parameters from one model on the same data (i.e., the same sample size), this dependence is of no importance here.

A, given actor A showed category i at the previous time point, and thus here the odds are ratios of conditional probabilities. An interaction effect, for example $\beta_{a_k^0 a_i^{-1} b_j^{-1}}$, can only be interpreted together with both main effects $\beta_{a_k^0 a_i^{-1}}$ and $\beta_{a_k^0 b_j^{-1}}$. The overall increase of the log odds of an observation k against K are then given by the sum of these three parameters, that is,

$$\beta_{a_k^0 a_i^{-1}} + \beta_{a_k^0 b_j^{-1}} + \beta_{a_k^0 a_i^{-1} b_j^{-1}}$$

gives the log-odds of behavioral category k against behavioral category K given actor A was in category i at the previous time point and actor B was in category j at the previous time point.

Software

In most popular packages for statistical data analysis, it is a problem to fit multinomial logit models for incomplete tables. Kroonenberg and Verbeek (1990) indicated they had great problems specifying the required models in the then current versions of SAS, BMDP, and SPSS. In the weighted least squares procedure of SAS the analysis without weights or structural zeros can be done without problems. The analyses including structural zeros or weights are a problem, however. As a package geared towards the analysis described, *ℓEM*[2] (Vermunt, 1996, 1997a, 1997b) overcomes such problems and can fit all models of interest. *ℓEM* is a program for the analysis of categorical data. It is possible to fit a wide range of models such as a variety of log-linear models, latent class models, correspondence analysis, and event history models. *ℓEM* is a quite general program, since the user can provide different kinds of restrictions on the models. For basic log-linear models and multinomial logit models the input file consists of a number of commands like the number of variables, the number of categories per variable, the model one wants to fit in log-linear notation, and the data. In Appendix A an annotated input file is presented that was used for one of our analyses. For our data with structural zeros, a weight vector can be specified for the margins which include cells which are set at missing. Zero starting values should be given to parameters concerning these margins. The other starting values can be set to any constant (usually 1) not equal to zero. *ℓEM* eliminates parameters in the correct places. The computation of the *df* is not automated yet. The output reports the number of cells minus the number of fitted parameters as the number of degrees of freedom. However, the reference guide tells us: 'Generally, we can correct the number of degrees of freedom by subtracting

[2] The program can be obtained from: http://www.kub.nl/faculteiten/fsw/organisatie/departementen/mto/software2.html

the number of fitted zeros and adding the number of non-identifiable parameters to the reported number of degrees of freedom' (Vermunt, 1997b, p. 75). This adjustment does provide the correct number, and both are given in the output. After adjusting the degrees of freedom, the *BIC*-statistic should also be adjusted, since it is based on the unadjusted number of degrees of freedom.

In event sampling the weights are on level four, that is, the weights are defined in terms of all four variables. In actor-specific event sampling the weights are on level 2, that is, on the level of a two-way interaction of child's previous behavior and child's current behavior. Consequently, in the first case (event sampling) it is impossible to define interactions on higher levels than the level of the weights. However, in the second case (actor-specific event sampling) it is possible to define interactions on a higher level than the weights. In that case a number of problems occur. It turns out that standard errors are not deleted in the right places. Moreover, the effects do not add up to zero. These problems occur because the program uses an internally defined design matrix.[3] It is possible that parameters are not identified. These problems do not affect the goodness-of-fit statistics (except for the *BIC*-statistic and the degrees of freedom, see above, but these values can be adjusted). Because these problems do not affect the goodness-of-fit statistics, first a model can be selected and then the parameter estimation problems can be dealt with. After a model is selected, a design matrix should be defined for the interactions in which the weighted cells occur. For example, if the weights are defined on the level of interaction between variable *A* and *C*, and the selected model includes this interaction a design matrix for this interaction effect should be defined. If an interaction effect is specified between *A*, *B*, and *C* a design matrix for this second-order interaction should be defined as well. Once the design matrix is specified it does not really matter whether we use the multinomial response the model of Equation 1 or multinomial logit the model of Equation 2. That is, for every row of our design matrix a parameter is estimated, and from these the parameters for both models can be derived. The program only gives standard errors for the estimated parameters, and not for the parameters derived from these estimates.

[3] A design matrix is used to define a priori contrasts. Design matrices are often dummy coded like in the regression analysis approach to analysis of variance. Probably the best known design matrices are contrast matrices in repeated measurement analysis (with for example Simple contrasts, Repeated contrasts, Difference contrasts, Helmert contrasts, end Polynomial contrasts).

Results

In this section the multivariate multinomial logit models will be used to answer the main research questions about play therapy data. Thus first the nature of all continuity and discontinuity transitions together will be examined and an adequate model for the complete data is derived. Then the questions about the relative influence of the therapist and child on the changes in behavior in the event sampling and actor-specific event sampling is answered.

Time Sampling

The four-way table of therapist and child's behavior at Lag 1 and Lag 0 (Table 2) will be analyzed. The variables at Lag 1 are denoted by *A* for the therapist and *B* for the child, the variables at Lag 0 are denoted *C* for the therapist and *D* for the child. *A* and *B* are the explanatory variables and *C* and *D* the response variables which we want to predict from the previous time point.

Since the influence of the explanatory variables on the response variables should be assessed, model [*AB*][*CD*] is used as baseline model, that is, the model of independence between the explanatory and response variables but saturated within each set. This is the standard approach in prediction analysis (see Von Eye, Brandstatter, & Rovine, 1993, 1998). Also in multivariate regression models and analysis of variance models the covariances between the explanatory and response variables are usually taken into account. As indicated in the section entitled "Model Selection, Evaluation, and Interpretation," only hierarchical models were fitted, moreover only a cross-dependence term of the therapist on the child was included if the auto-dependence term of the child was also included in the model (and vice versa).

Model Selection

The 27 models of interest that were fitted to the four-way table are given in Table 3. In all models the interaction between therapist and child at Lag 1 (Variables *A* and *B*) and that at Lag 0 (Variables *C* and *D*) were present. The model with only these two interactions was our base model (Model 1). All subsequent models introduce other effects, if a cross-dependence effect was present, an auto-dependence effect was also included. Only hierarchical models are used to be sure that when an interaction is present also the main effects are in the model. In Table 3 the models are shown using

Table 3
Goodness-of-Fit Measures for Models without Weights

Model	*df*	X^2	*LR*	*D*	*BIC*	R^2
1.	225	124222	87914	.4491	85399	.0000
2.[AC][BD]	207	7984	6731	.0981	4418	.1831
3.[AC][BD][AD]	198	6909	6264	.0945	4052	.1828
4.[AC][BD][BC]	198	5815	5390	.0907	3177	.1866
5.[AC][BD][AD][BC]	189	4282	4300	.0878	2188	.1888
6.[ABC]	180	60177	45932	.3217	43920	.0742
7.[ABC][BD]	171	4098	3725	.0692	1814	.1813
8.[ABC][BD][AD]	162	2605	2602	.0604	791	.1828
9.[ABD]	180	29669	22685	.2222	20674	.1280
10.[ABD][AC]	171	5537	4797	.0787	2886	.1833
11.[ABD][AC][BC]	162	2795	2771	.0675	960	.1886
12.[ABC][ABD]	**135**	**1471**	**1339**	**.0393**	**–169**	**.1843**
13.[ACD][BD]	171	5699	4938	.0766	3027	.1758
14.[ACD][BD][BC]	162	3043	2960	.0628	1150	.1815
15.[BCD][AC]	171	4254	3927	.0761	2016	.1858
16.[BCD][AC][AD]	162	2834	2792	.0679	981	.1883
17.[ACD][BCD]	135	1841	1676	.0445	167	.1819
18.[ABC][BCD]	144	3091	2262	.0399	653	.1811
19.[ABD][ACD]	144	4851	3471	.0529	1862	.1773
20.[ABC][ACD][BD]	135	2169	2225	.0588	716	.1810
21.[ABD][BCD][AC]	135	2053	2035	.0577	526	.1885
22.[ABC][BCD][AD]	135	1531	1367	.0373	–140	.1836
23.[ABD][ACD][BC]	135	1865	1740	.0436	231	.1830
24.[ABC][ABD][ACD]	108	1143	1049	.0339	–157	.1828
25.[ABC][ABD][BCD]	108	844	807	.0299	–399	.1846
26.[ABC][ACD][BCD]	108	1151	1044	.0332	–162	.1819
27.[ABD][ACD][BCD]	108	1171	1116	.0334	–343	.1831
28.[ABC][ABD][ACD][BCD]	81	578	562	.0308	–343	.1832

Note. *A* and *C* are therapist behavior at Lag 1 and Lag 0, respectively; *B* and *D* are child behavior at Lag 1 and Lag 0, respectively.

the usual notation for hierarchical log-linear models. Screening the fit statistics we see that no one model fitted the data according to the two traditional chi-square distributed statistics. However, there were large differences in both X^2 and LR between the models in the table. To get anywhere near reasonable fit statistics, all auto-dependence and cross-dependence effects are needed. If at least 95% of the predictions should be classified correctly, the dissimilarity index (D) points towards Model 12, 17, 18 or 22-28. For all these models, except 17, 18, and 23, the BIC is negative, eliminating these three from the list of candidate models. Finally, models 24-28 use many more parameters than Models 12 and 22, so that the choice had to be made between the latter models.

In order to assist model selection, we constructed a graph for the standardized versions of all our fit statistics (Figure 1) using $-R^2$ because this statistic is the only goodness-of-fit where the others are badness-of-fit statistics. From the figure it can be seen that all fit statistics convey the same information about the fit of the models. The most important conclusion from this figure is that models 1, 6, and 9, fit very badly. In Figure 2 these models

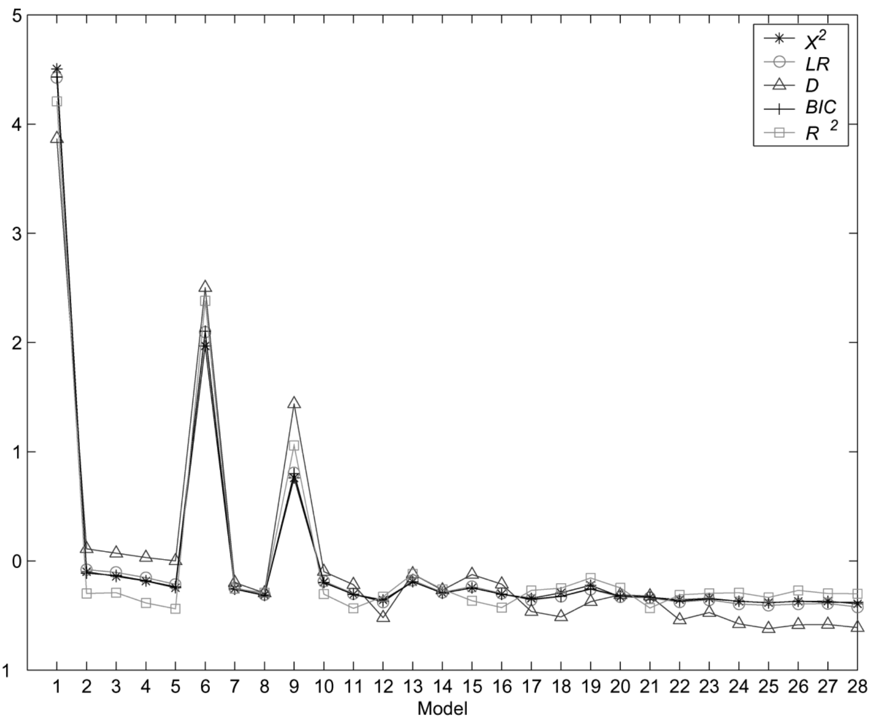

Figure 1
Standardized Fit Statistics for all Models of Table 3

are deleted from the graph to get a better view of the fit of the remaining models. Moreover, the models are sorted with respect to the degrees of freedom and model with equal *df* are ordered according to increasing values of X^2 statistic (models with equal *df* are between vertical lines). In Figure 2 we see that the fit increases with increasing *df*, but that the sizes of the increases level off at the end. The fit increase between block (12, 22, 23, 17, 21, 20) and block (25, 24, 26, 28) is relatively small compared to the fit increase between the earlier blocks. Therefore, it seems good to select a model from the former block. Again a choice has to be made between Model 12 and 22, but no clear preference emerges between the two, so that we should look at parsimony and interpretability.

Although models 12 and 22 have the same number of *df*, the number of effects to interpret is much larger in Model 22. This can be seen from the description of models in Table 3, where for Model 12 two three-way effects are given (*ABC* and *ABD*) and for Model 22 two three-way effects (*ABC* and *BCD*) plus a two-way effect (*AD*). For interpretational parsimony we should choose for Model 12 instead of Model 22. In Model 12 we have no actor-

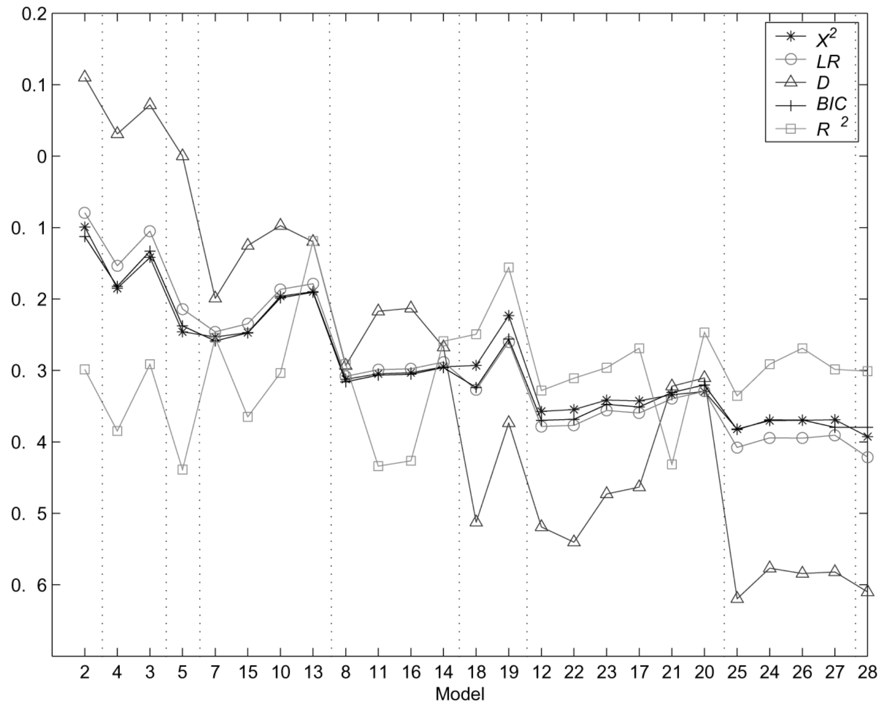

Figure 2
Standardized Fit Statistics for all Models of Table 3 except 1, 6, 9

dependence effects, only two interaction-dependence effects, which is nicely symmetric. In Model 22 we have one actor-dependence effect (the interaction is dependent on the child's previous behavior), one interaction-dependence effect (an effect of the interaction of both actors on the future behavior of the therapist), and a cross-dependence effect of the child on previous behavior of the therapist.

We conclude that Model 12 is the best and simplest model of the two. It has an interaction-dependence effect on both the therapist and the child, and its *BIC*-statistic is negative, indicating a good fit. Also the R_C^2 statistic is reasonably large, 18.4% of the dispersion is accounted for, where no one model reaches 19%.

The selected model is shown graphically in Figure 3. There are two boxes, one for the explanatory variables and one for the response variables. In each box the margins of the variables are fixed by design (there is an interaction between the two actors at Lag 0 and Lag 1). Solid dots represent the variables, and an open dot represents an interaction point of the lines. Vectors denote the effects. In the selected model the interaction between the therapist and the child has an effect on the behavior of the child and the interaction between therapist and child has an effect on the therapist.

Similar graphs were proposed earlier by Goodman (1973). Other graphical models, which are actually called 'Graphical Models', have been researched intensively (Cox & Wermuth, 1996; Edwards, 1995; Whittaker, 1990). However, these models do not contain interaction points. The standard

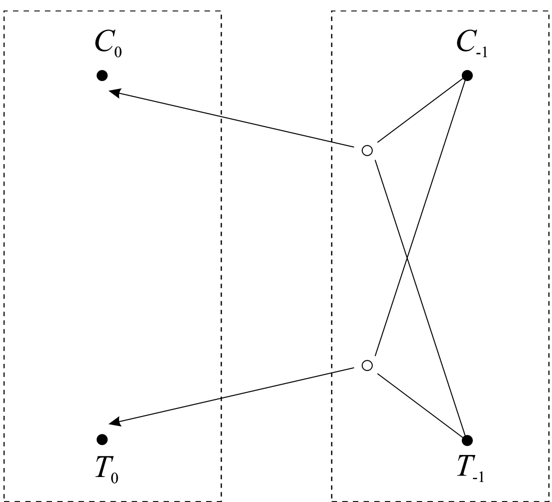

Figure 3
Selected Model Without Weights

representation of graphical models is followed, in which the response variables are on the left-hand side, and the explanatory variables on the right-hand side.

Graphs as presented in Figure 3 are very useful because they represent rather complex models in a very detailed but easy manner. These graphs visualize dependencies between variables, and therefore the interpretation of obtained results is more straightforward.

Model Interpretation

In Table 4 the results of the analysis with Model 12 are shown. The columns pertain to the future behavior, the rows to the past behavior. The columns give the log-odds of a specific category against the baseline category Imagery play, for example in the column N vs. I the log odds of Non play compared to Imagery play are given. The left part is for the future child behavior, the right part for the future therapist behavior. The rows are ordered using blocks. The first block gives the effects of the Child previous behavior on the actors future behavior. The second block gives the effects of the therapists past behavior on the actors future behavior, the last four blocks give the effects of the interaction of child and therapist on the future behavior of each actor. For reasons of space no standard errors are presented. After the description of each effect the Wald statistic for that effect is given. This statistic is a description of the total strength of the effect. Where each parameter together with its standard error shows whether the parameter is different from zero, the Wald statistic does actually the same but for a group of parameters. The Wald statistics is used to evaluate the dominance concepts. Comparing the cross-dependency effects of child on therapist ($W = 1035$, $df = 9$) and vice versa ($W = 1330$, $df = 9$), it can be concluded that they only differ by a small amount (ratio is 1.28): none of the partners is directly dominant over the other. Since the effects of both the child and the therapist on the future interaction are not in the model, that is, they are equal to zero, there is no indirect dominance either.

The parameter estimates for the main terms show a strong *continuity* effect as the largest entries in the table correspond to repeated behavior. For instance, given a child does not play at t_{-1} (C_N^{-1}), the probability that he will continue to do so $Pr(C_N^0|C_N^{-1})$ versus the probability that he will be involved in Imagery play at t_0, $Pr(C_I^0|C_N^{-1})$, is $\exp(1.542) = 4.70$. In other words, given Non play at t_{-1} the child is 5 times as likely to continue to do so, as it is to switch to Imagery play. After Imagery play at t_{-1} all odds are in favor of Imagery play again, indicated by the minus signs. These are the highest values in the table, that is the strongest effects. This strong continuity is characteristic of these data and partly due to the scoring of every five seconds which period turns out to be shorter than the duration of most behavior.

Table 4
Parameter Estimates of Selected Model Without Weights

	Child			Therapist		
	N vs. I	P vs. I	F vs. I	N vs. I	P vs. I	F vs. I
	Auto-dependence, W_9 = 14318			Cross-dependence, W_9 = 1035		
C_{-1}						
Non play	1.542	.539	.287	.472	.548	.391
Play preparation	.395	1.585	.374	.310	.499	.084
Functional play	.533	.542	2.260	.190	.109	.705
Imagery play	–2.470	–2.666	–2.921	–.973	–1.156	–1.179
	Cross-dependence, W_9= 1330			Auto-dependence, W_9 = 12098		
T_{-1}						
Non play	.249	.159	.078	.827	.038	–.194
Play preparation	.208	.483	.071	.360	1.635	.324
Functional play	.492	.339	.914	.205	.256	1.856
Imagery play	–.949	–.981	–1.063	–1.392	–1.156	–1.986
	W_{27}= 1206, Interaction-dependence, W_{27} = 1370					
$T_{-1} \times C_{-1}$						
Non play						
Non play	.669	.225	.321	.551	.233	.280
Play preparation	–.267	–.175	–.229	–.212	.219	.046
Functional play	–.042	–.048	.068	–.007	.093	.370
Imagery play	–.360	–.003	–.160	–.332	–.545	–.696
Play preparation						
Non play	.000	.132	–.052	–.072	–.107	.053
Play preparation	.097	–.226	–.140	.029	–.306	–.150
Functional play	.180	–.004	.292	.173	–.026	.223
Imagery play	–.277	.098	–.100	–.130	.442	–.126
Functional play						
Non play	.083	.325	.553	–.008	.192	–.010
Play preparation	.215	.343	.443	.165	.046	.183
Functional play	–.031	.115	–.488	.065	.147	–.326
Imagery play	–.267	–.783	–.508	–.223	–.384	.153
Imagery play						
Non play	–.752	–.683	–.822	–.471	–.318	–.322
Play preparation	–.045	.059	–.074	.018	.045	–.079
Functional play	–.107	–.063	.127	–.232	–.214	–.268
Imagery play	.904	.687	.768	.685	.487	.669

Note. Log-odds of .693 indicate the numerator is twice as likely as the denominator [exp(.693) = 2]. Similarly, 1.099 ⇒ 3×, 1.386 ⇒ 4×, 1.609 ⇒ 5×, and 2.303 ⇒ 10×. Minus signs reverse the probability: The denominator is more likely than the numerator.

The behavior of therapist and child at t_{-1} show some interesting interactions with behavior at t_0. In evaluating the size of such interactions, one should realize that they only provide a contribution to the overall log-odds over and above the main terms included in the model. The interaction should therefore be interpreted as an additional emphasis or de-emphasis of the main terms.

In general, when both child and therapist show Imagery play at t_{-1}, the probability reduces that one of them shows this mode of play again at t_0 compared to the effect of only the two main terms (.90, .69, .77 for Non Play, Preparatory, and Functional play versus Imagery play of the Child, respectively; .69, .49, and .67 for the same therapist behaviors at t_0). This holds true also for the other categories, except for Non play. This indicates that adding the log-odds when both partners are in the same category overstates the continuity of the joint behavior with respect to Imagery play. The log-odds when both partners are in Imagery play at t_{-1} are the largest interactions, mediating the strong continuity suggested by the main terms alluded to above. However, given Non play of both partners at t_{-1}, the probability that they show Non play again rather than Imagery play at t_0 is increased with the log-odds being larger for the child behavior (.67) than for the therapist behavior (.55). Thus even though in general the odds are more or less even for any behavior of the child with respect to Imagery play when the therapist is in Non play (.25, .16, .08), the log-odds increase towards the other behaviors when in addition at t_{-1} the child is also not playing (.67, .23, .32).

A final indication that the child follows the therapist comes from the size of the log-odds when at t_{-1} the therapist shows Imagery play and the child Non play. Then the probability is higher at t_0 that the child also shows Imagery play as well versus any other category (–.75, –.68, –.82). Thus at t_0 the probability increases that they are in Imagery play together.

Event Sampling

Model Selection

The same models as those in the section entitled "Model Selection" are analyzed. Here however in the model the categories on the main diagonal of Table 2 are weighted by zeros. The results are shown in Table 5. Only the last model with the four-way interaction deleted fits the data when solely relying on the chi-square distributed statistics and this model explains 12.6% of the dispersion. However, the model with only main effects (Model 5) explains 12.4% of the dispersion, and it has a negative *BIC*-value, indicating the model fits relatively better than the saturated model. The dissimilarity index for this model has the value .05, so about 5% of the observations have

Table 5
Goodness-of-Fit Measures for Models with Weights for Event Sampling

Model	*df*	X^2	*LR*	*D*	*BIC*	R^2
1.	209	39865	32876	.3694	30661	.0000
2.[AC][BD]	191	6012	4458	.0993	2434	.1090
3.[AC][BD][AD]	182	4725	3950	.1014	2021	.1119
4.[AC][BD][BC]	182	3440	2988	.0856	1059	.1173
5.[AC][BD][AD][BC]	**173**	**1021**	**1040**	**.0515**	**–793**	**.1241**
6.[ABC]	164	23778	18778	.2645	17040	.0398
7.[ABC][BD]	155	2670	2469	.0805	826	.1218
8.[ABC][BD][AD]	146	734	751	.0412	–796	.1261
9.[ABD]	164	12502	9596	.1689	7858	.0947
10.[ABD][AC]	155	4003	3251	.0850	1608	.1135
11.[ABD][AC][BC]	146	726	733	.0404	–814	.1245
12.[ABC][ABD]	119	513	511	.0318	–749	.1266
13.[ACD][BD]	155	4175	3592	.0976	1949	.1132
14.[ACD][BD][BC]	146	886	899	.0463	–648	.1239
15.[BCD][AC]	155	2768	2425	.0754	782	.1162
16.[BCD][AC][AD]	146	814	816	.0438	–731	.1241
17.[ACD][BCD]	119	709	711	.0397	–549	.1239
18.[ABC][BCD]	128	2040	1610	.0500	253	.1198
19.[ABD][ACD]	128	3433	2544	.0658	1187	.1151
20.[ABC][ACD][BD]	119	555	573	.0344	–687	.1266
21.[ABD][BCD][AC]	119	538	537	.0334	–723	.1255
22.[ABC][BCD][AD]	119	574	571	.0322	–689	.1259
23.[ABD][ACD][BC]	119	546	646	.0380	–614	.1245
24.[ABC][ABD][ACD]	92	339	340	.0227	–634	.1268
25.[ABC][ABD][BCD]	92	303	303	.0229	–671	.1268
26.[ABC][ACD][BCD]	92	387	386	.0246	–588	.1259
27.[ABD][ACD][BCD]	92	446	442	.0298	–532	.1250
28.[ABC][ABD][ACD][BCD]	65	70	70	.0086	–618	.1261

Note. *A* and *C* are therapist behavior at Lag 1 and Lag 0, respectively; *B* and *D* are child behavior at Lag 1 and Lag 0, respectively.

to be moved to obtain a perfect fit. From this screening, it can be concluded that Model 5 is the most parsimonious model that fits the data well enough.

Standardizing our statistics and plotting them gives Figures 4 and, more detailed (Models 1, 6 and 9 left out), Figure 5. Especially from Figure 5, which is ordered from most parsimonious to least, it is clear that Model 5 shows a large increase in fit compared to models with approximately equal *df*. The larger models (i.e. with less *df*) show hardly an increase in fit in most statistics (for this series the dissimilarity index seems to be somewhat more outspoken). Since Model 5 is much simpler and more parsimonious than Models 11-28, it is chosen.

A graphical representation of the model is shown in Figure 6. There are no interaction points anymore, only direct effects of the therapist to itself and the child and of the child to the therapist and itself.

Model Interpretation

The parameter estimates of Model 5 are given in Table 6. Here the Wald statistic for the influence of the child on the therapist is larger than the Wald

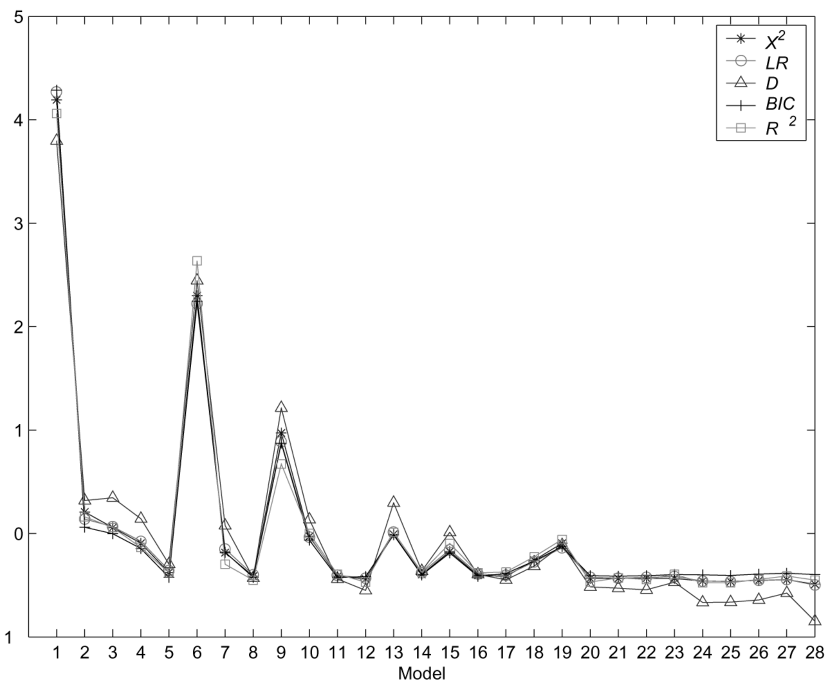

Figure 4
Standardized Fit Statistics for all Models of Table 5

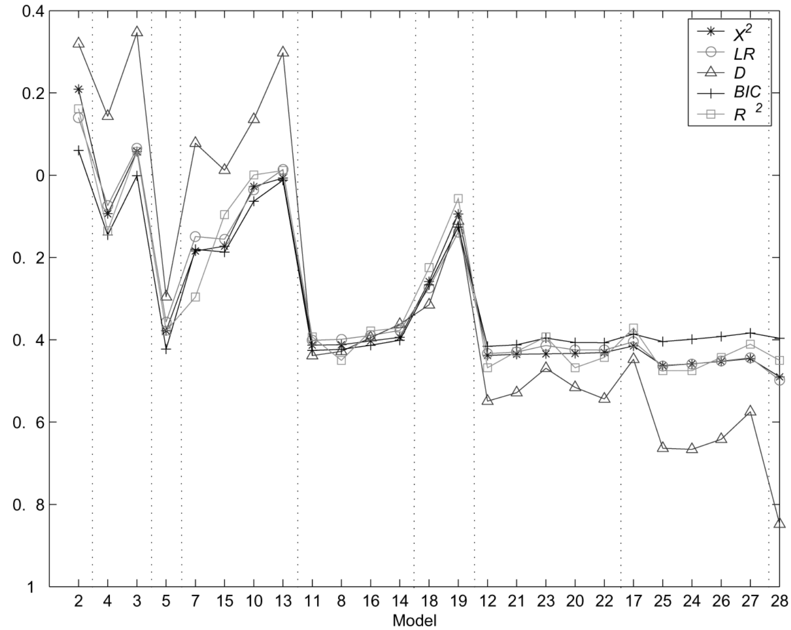

Figure 5
Standardized Fit Statistics for all Models of Table 5 except 1, 6, 9

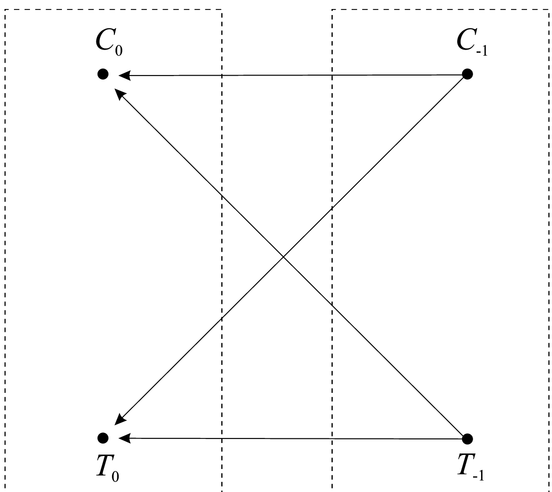

Figure 6
Selected Model with Weights for (actor-specific) Event Sampling

statistic for the influence of the therapist on the child, but again the difference is relatively small (ratio = 1.47). So neither the therapist nor the child are directly or indirectly dominant.

Compared to the time-sampling analysis the interaction-dependence effects were not needed anymore. This indicates that the interaction-dependence effects mainly dealt with the continuity in the data. In event-sampling the simultaneous continuity of both actors does not play a role anymore, and the interaction-dependence effects were not needed to represent the data.

Similarly to the analysis without weights, the entries corresponding to no change of behavior are huge compared to the other parameter estimates. For example for the child the values corresponding to no change are.94, 1.44, 2.13, –2.05, –2.29, and –2.55. Minus signs are found after Imagery play, indicating that once Imagery play is initiated, it continues. It should be noted here that if the child continues to play in one mode, the therapist changes of behavior, since the cells that relate to no change of the dyad are treated as missing. All other behaviors lead to play modes other then Imagery play. The values after Non play of the therapist are the smallest, indicating the therapist best shows Non play or Imagery play to draw the child to Imagery play.

Comparing the *auto-dependence* effects of child and therapist, the parameter estimates of the child are somewhat higher in absolute value than

Table 6
Parameter Estimates of Selected Model with Weights for Event Sampling

	Child			Therapist		
	N vs. I	P vs. I	F vs. I	N vs. I	P vs. I	F vs. I
	Auto-dependence, $W_9 = 12428$			Cross-dependence, $W_9 = 2919$		
C_{-1}						
Non play	.942	.307	.110	.372	.544	.345
Play preparation	.431	1.439	.313	.429	.686	.145
Functional play	.674	.549	2.130	.253	.122	.909
Imagery play	–2.046	–2.294	–2.552	–1.053	–1.352	–1.399
	Cross-dependence, $W_9 = 1980$			Auto-dependence, $W_9 = 4822$		
T_{-1}						
Non play	.111	.163	.066	.484	.045	–.185
Play preparation	.440	.568	.186	.363	1.329	.259
Functional play	.433	.361	.906	.247	.294	1.684
Imagery play	–.984	–1.092	–1.158	–1.093	–1.668	–1.758

the parameter estimates for the therapist, indicating the child shows greater auto-dependence. The *cross-dependence* effects for child and therapist are rather similar, some effects are larger for the influence of the child on therapist, others are the reverse. A general pattern is that a particular behavior by one of the partners is followed by the same behavior of the other partner, as can been seen from the main diagonals of the subtables. The main exception is the influence of Non play of the therapist on the child's Non play [$Pr(C^0_N|T^{-1}_N)$] over Imagery play [$Pr(C^0_I|T^{-1}_N)$] which parameter value is .11, which is low and seems in accordance with the situation that the therapist has a purpose and wants to achieve something, that is, Imagery play, while the child does not have such a goal. Moreover, the Non play category of the therapist is primarily 'simple attention' indicating that the therapist leaves the initiative with the child.

A further noteworthy pattern is that Non play of the child at t_{-1} enhances the probability of all other therapist activities with respect to Imagery play (0.37, 0.54, 0.35), which illustrates that Imagery play generally does not start at Non play of the child, and that it needs to be introduced via other play activities. The reverse is not true, as for the therapist the odds are all much smaller and almost even (0.11, 0.16, 0.07), which suggests that Non play of the therapist has less influence on the child's next activity with respect to Imagery play. Note furthermore, that nearly all log-odds, accept those for Imagery play at t_{-1} are positive indicating that the transitions from any other activity to Imagery play are infrequent. However, the log-odds for Imagery play to Imagery play are the largest in the table, indicating that once it is established it tends to continue.

Actor-Specific Event Sampling

Model Selection

Again the same series of models is applied. The goodness-of-fit statistics are shown in Table 7. The simplest model with a negative *BIC*-value is again the model with all main effects, that is, the model with both *auto-dependence* and *cross-dependence* effects for both therapist and child, and therefore Figure 6 also displays this model. The percentage of dispersion accounted for is low 7.2%, but even the model with all three-way interactions (Model 30) does not account for more then 7.4%. The dissimilarity index is .065 reasonably small. Also the chi-square statistics show a large decrease compared to the simpler models and some higher models. Figures 7 and 8 give the graphed standardized statistics (in 8, Models 1 and 9 are left out, and the models are ordered with respect to *df*). Like in

Table 7
Goodness-of-Fit Measures for Models with Weights for Actor-Specific Event Sampling

Model	*df*	X^2	*LR*	*D*	*BIC*	R^2
1.	161	13753	11265	.2977	9664	.0000
2.[AC][BD]	146	4611	3744	.1502	2292	.0472
3.[AC][BD][AD]	137	2679	2367	.1169	1005	.0616
4.[AC][BD][BC]	137	1974	1869	.1050	506	.0591
5.[AC][BD][AD][BC]	**128**	**682**	**681**	**.0657**	**-591**	**.0724**
6.[ABC]	116	1721	1514	.0844	360	.0614
7.[ABC][BD]	111	1646	1456	.0818	352	.0616
8.[ABC][BD][AD]	101	428	433	.0484	-571	.0745
9.[ABD]	132	9364	8161	.2515	6849	.0192
10.[ABD][AC]	123	2528	2232	.1080	1009	.0623
11.[ABD][AC][BC]	113	549	548	.0567	-575	.0733
12.[ABC][ABD]	87	314	315	.0404	-549	.0754
13.[ACD][BD]	111	2555	2135	.1032	1031	.0607
14.[ACD][BD][BC]	102	579	578	.0610	-436	.0715
15.[BCD][AC]	122	1891	1780	.1006	567	.0596
16.[BCD][AC][AD]	114	593	599	.0606	-534	.0729
17.[ACD][BCD]	87	484	487	.0560	-378	.0722
18.[ABC][BCD]	95	1600	1367	.0769	422	.0625
19.[ABD][ACD]	95	2389	2001	.0947	1057	.0614
20.[ABC][ACD][BD]	75	203	207	.0313	-539	.0729
21.[ABD][BCD][AC]	99	507	506	.0548	-478	.0735
22.[ABC][BCD][AD]	86	356	361	.0417	-494	.0749
23.[ABD][ACD][BC]	87	438	442	.0510	-423	.0724
24.[ABC][ABD][ACD]	60	87	89	.0180	-507	.0735
25.[ABC][ABD][BCD]	72	280	280	.0383	-436	.0754
26.[ABC][ACD][BCD]	60	120	121	.0206	-476	.0733
27.[ABD][ACD][BCD]	71	396	394	.0477	-312	.0729
28.[ABC][ABD][ACD][BCD]	17+[a]	50	52	.0122	-117	.0740

Note. *A* and *C* are therapist behavior at Lag 1 and Lag 0, respectively; *B* and *D* are child behavior at Lag 1 and Lag 0, respectively.
[a]If the number of parameters is larger than 150 the program does not give any identification information. So for the number of parameters the boundary parameters have to be added and the BIC statistic has to be changed consequently.

the event-sampling section, Model 5 is clearly better than comparable-*df* models, and the fit statistics have the same size as those of the less parsimonious models (16 to 28).

Model Interpretation

In the case that the child's behavior is weighted but not that of the therapist, an analysis to see which of the partners is dominant is impossible. So the interest in this analysis can only be in the parameter estimates of the model given in Table 8.

In this model the diagonal cells for the child are treated as structural zeros, the parameters referring to these cells do not exist (these are left blank in Table 8). The three differences against Imagery play do not tell the whole story anymore and some other differences are needed to get all the information. This is done in the first block of Table 8. The structure is the

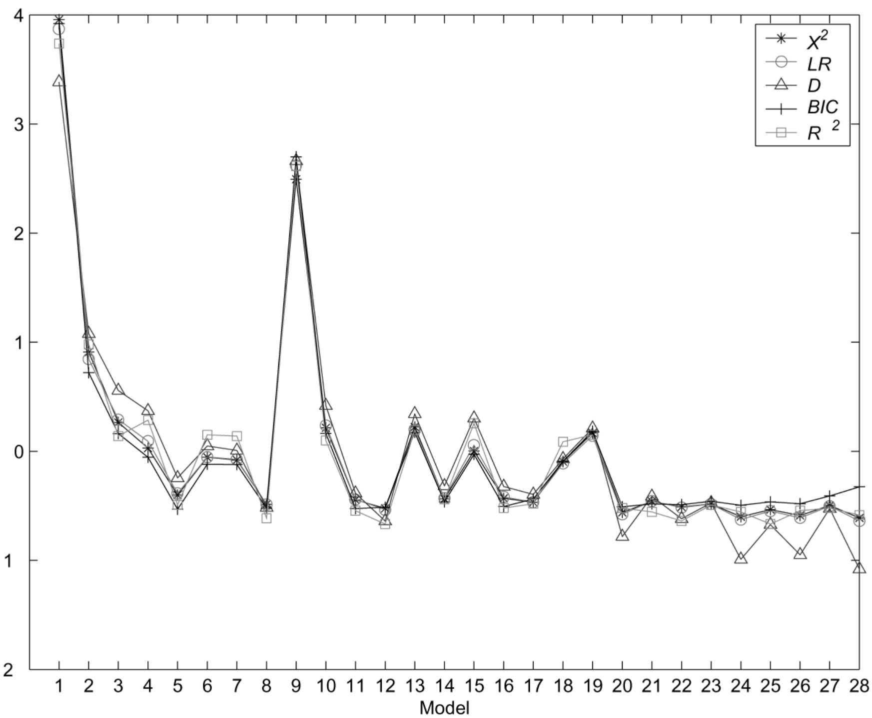

Figure 7
Standardized Fit Statistics for all Models of Table 7

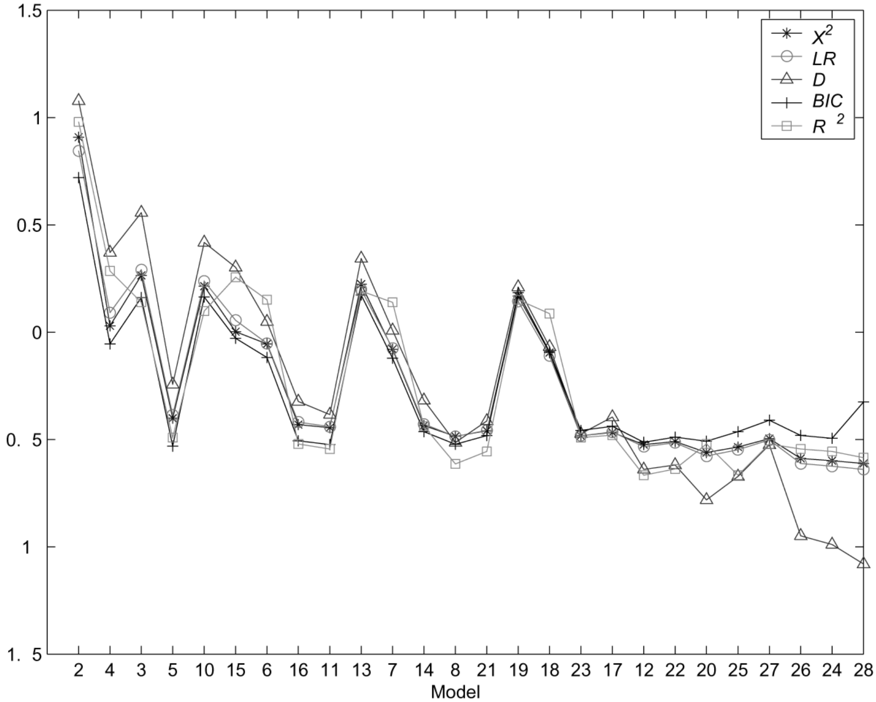

Figure 8
Standardized Fit Statistics for all Models of Table 7 except 1, 9

same as in previous tables, in the rows are the past behaviors in the columns the future behaviors, again in the form of log-odds. Again N vs. I denotes the log-odds of Non play versus Imagery play. For the other effects, the differences against Imagery play are still sufficient. In the first block, the interest is focused on changes of the child and the parameters are much smaller compared to the previous analyses. Moreover, the pattern of positives and negatives is not so clear anymore. After Non play of the child the odds are in favor of Imagery play compared to Play preparation and Functional play (–.082 and –.049 respectively), although the effects are small. After Play preparation of the child, the odds are in favor of both Imagery play and Functional play compared to Non play, where the effect of Functional play is somewhat stronger. Furthermore, after Functional play the odds are against Imagery play. After Imagery play the odds are always against Functional play, supporting the idea that Functional play and Imagery play are two end points in a play sequence.

Table 8
Pameter Estimates of Selected Model with Weights for Actor-Specific Event Sampling

	Child/Therapist					
	N vs. I	P vs. I	F vs. I	N vs. F	P vs. F	F vs. P
	Auto-dependence, $W_5 = 41$					
C_{-1}						
Non play		–.082	–.048		–.034	
Play preparation	–.150		.041	–.191		
Functional play	.142	.097				.045
Imagery play				.124	.095	.030
	Cross-dependence, $W_9 = 1643$					
C_{-1}						
Non play	.366	.542	.305			
Play preparation	.494	.764	.247			
Functional play	.258	.032	1.161			
Imagery play	–1.117	–1.337	–1.714			
	Cross-dependence, $W_9 = 1175$					
T_{-1}						
Non play	.139	.155	.010			
Play preparation	.294	.613	-.028			
Functional play	.504	.484	1.253			
Imagery play	–.938	–1.251	–1.235			
	Auto-dependence, $W_9 = 3868$					
T_{-1}						
Non play	.522	-.031	-.162			
Play preparation	.272	1.246	.330			
Functional play	.216	.365	1.483			
Imagery play	–1.010	–1.580	–1.651			

The other three blocks give the usual pattern of continuity of behavior. A noteworthy change compared to earlier results is that after Play preparation of the therapist the odds are in favor of Imagery play compared to Functional play (–.028). Although the magnitude of this number is small it is important that here the sign has changed. Furthermore, we see again that the therapist best shows Non play (the odds are about even for the child's behavior) or Imagery play (the child follows).

Discussion

In this article multivariate multinomial logit models were applied to dyadic sequential data, even though dyadic sequential data analysis most often are analyzed with log-linear models. However, the latter approach has a major drawback, in that the interpretation of the parameters is rather awkward. Moreover, often only a small number of models are fitted and primarily the residuals from these models are examined rather than their parameters themselves. In our opinion, one should fit models and interpret the parameters of the well-fitting ones rather than examining and interpreting their residuals. To examine well-fitting models, one could fit them with weighted least squares procedures as discussed by Budescu (1984), but this approach is only available in the program SAS, while log-linear modeling is available in many more mainstream statistical packages. The multinomial logit model is a simple reparametrization of the log-linear model, but one that allows for the more easily interpreted odds. However, the use of multinomial logit models with structural zeros is infeasible in main stream statistical packages.

Handling of Structural Zeros

We showed how the reparametrization is done and how parameter estimates can be interpreted. Problems in fitting multivariate multinomial logit models occur when there are structural zeros or when zero weights are attached to certain cells of our contingency table. Such problems were discussed and a procedure was devised to obtain valid results. It turned out that the problems with structural zeros in multinomial logit models are very much the same as those encountered using the weighted least squares approach to the analysis of contingency tables (compare Wickens, 1989, chapter 12). In either case correct solutions are obtained by specifying a design matrix.

Modeling Individual Differences

Although we treated all dyads together, in principle it is possible to study individual differences between dyads. In our study, for example, we have therapy and non-therapy children. A variable G could be made which indicates whether the dyad involves a therapy or non-therapy child. The basic model to study is $[GAB][GCD]$ and we can detect whether there are differences between the therapy and non-therapy group with respect to predicting of future behavior using essentially the same types of models. The number of possible models will increase considerably, but this should not be a argument to neglect individual differences. Theoretically, the variable G could refer to each dyad separately. In practice, however, one should realize that a variable G with two categories doubles the number of cells in the contingency table, and sparseness of the data becomes a real problem. When the variable G refers single dyads, the number of cells would, in our case, have to be multiplied with 117, and the resulting contingency table would be extremely sparse. In the application shown (with only 1 group) the number of cells is $4^4 = 256$. For each category of the variable G these 256 cells could be reasonably filled. Taking into account that most behavior is continue (i.e. behavior a_i is most often followed by a_i), the number of observations on a dyad should be very large.

Graphing the Data

In all models presented interpretation of results necessarily involved comparing many numbers. An alternative procedure is to analyze the data using models based on bilinear decompositions of the interaction parameters such as in correspondence analysis (Greenacre, 1984), or transforming interaction parameters to distances (De Rooij, 2001). In these procedures interactions are represented graphically. In the first case through projections, in the second case through distances. Such graphical representations can facilitate interpretation enormously, and interpretation of results can be done looking at a single graphical representation.

Modeling versus Testing Effects

Our approach to the analysis of dyadic sequential data is modeling of the whole data set, where Budescu (1984) more specifically tested hypotheses about (indirect) dominance. In the weighted least squares approach it is also possible to adopt the modeling approach, but for each model a design matrix has to be specified. In our procedure only design matrices have to be specified in the case of actor-specific event sampling. Moreover we can first select a model and afterwards make a design matrix for the selected model to obtain the correct parameter estimates. The interpretation of the weighted least squares approach is in terms of regression coefficients, where the interpretation of our approach is in terms of log-odds. Both are more attractive than the standard log-linear model parameters, but the choice between regression coefficients and log-odds must be made on personal grounds. In his paper Budescu (1984) gave explicit tests for (in)direct dominance. We could only give a descriptive way of assessing whether there exist dominance, since the two Wald statistics are dependent. In our examples this descriptive way was strong enough to conclude that neither the child nor the therapist was (in)directly dominant. The weighted least squares approach is available in the mainstream statistical package SAS, where our approach can be handled by using the program ℓEM, which is not part of a major commercial package, but it is freely available on the internet.

Goodness-of-fit

To select a model we used a number of goodness-of-fit statistics. If we look at Tables 3, 5 and 7 and corresponding Figures of z-transformed statistics (e.g., 1, 2, 4, 5, 7, 8) we see that all statistics point towards the same models. There are some small differences, but in general any statistic can be used to select a model. Our preference goes to the *BIC*-statistic since it has a clear reference point: If the statistic is smaller than zero, the model fits relatively better than the saturated model. For forecasting the dissimilarity index is important, since it gives the proportion of misclassification.

Modeling Other Types of Behavior

In this article we looked at changes of behavior of the dyad or of one of the actors. We did not look at the effect of duration of the previous behaviors on the changes made. Also in the therapy data duration plays an important role. As an example, we think of the therapist who keeps on showing Imagery play till the child follows his or her example. In that case, it is interesting to know how long the therapist needs to go on showing Imagery play before the child follows. Here the approach of Griffin and Gardner (1989) could be taken.

We often see that one of the actors has a long sequence of the same behavior, but that he or she for a very short period shows another kind of behavior and then returns to the same behavior as before. It would be interesting to see what would happen if we smoothed the sequence of behaviors. This, however, poses several difficulties concerning how to smooth the sequence of observations and requires a separate paper.

Results of the Three Analyses

The three analyses might be summarized as follows. In the time-sampling analysis the results pertain mainly to the stayers in the data. The two interaction effects had to be included to downweight the effects of the main effects. The second analysis pertains to the movers, that is the stayers are left out of the analysis and results pertain now to changes of the dyads. We found that it is difficult for the therapist to draw the child to Imagery play. The best the therapist can do is play Imagery and hope the child will follow. Otherwise, the therapist is best involved in Non play, since in that case the odds are about even for other kinds of play and Imagery play. In actor-specific event sampling the results compared to the event sampling do not change much, although we see that some nice changes of sign, that is some odds are in the other direction. For example, after Play preparation the odds are in favor of Imagery play against Functional play.

References

Agresti, A. (1990). *Categorical data analysis.* New York: John Wiley and Sons.

Agresti, A. (1996). *An introduction to categorical data.* New York: John Wiley and Sons.

Allison, P. & Liker, J. K. (1982). Analyzing sequential categorical data on dyadic interaction: A comment on Gottman. *Psychological Bulletin*, *91*, 393-403.

Bakeman, R. & Gottman, J. M. (1997). *Observing interaction: An introduction to sequential analysis.* Cambridge, UK: Cambridge University Press.

Bakeman, R. & Quera, V. (1995). Log-linear appraches to lag-sequential analysis when consecutive codes may and cannot repeat. *Psychological Bulletin*, *103*, 272-284.

Benes, K. M., Gutkin, T. B., & Kramer, J. J. (1995). Lag sequential analysis: Taking consultation communication research methodology to the movies. *School Psychology Review*, *24*, 694-708.

Budescu, D. V. (1984). Tests for lagged dominance in sequential dyadic interaction. *Psychological Bulletin*, *96*, 402-414.

Cox, D. R. & Wermuth, N. (1996). *Multivariate dependencies: Models, analysis, and interpretation.* London: Chapman and Hall.

De Rooij, M. (2001). *Distance models for the analysis of transition frequencies.* Unpublished doctoral dissertation, Leiden University.

Edwards, D. (1995). *Introduction to graphical modelling.* New-York: Springer-Verlag.

Fienberg, S. E. (1980). *The analysis of cross-classifed categorical data.* Cambridge, MA: MIT Press.

Goodman, L. A. (1968). The analysis of cross-clasifed data: Independence, quasiindependence, and interactions in contingency tables with and without missing entries. *Journal of the American Statistical Association*, *63*, 1091-1131.

Goodman, L. A. (1973). The analysis of multidimensional contingency tables when some variables are posterior to others: A modifed path analysis approach. *Biometrika*, *60*, 179-192.

Gottman, J. M. (1979a). Detecting cyclicity in social interaction. *Psychological Bulletin*, *86*, 338-348.

Gottman, J. M. (1979b). *Marital interaction: Experimental investigations.* New York: Academic Press.

Gottman, J. M. (1981). *Time-series analysis: A comprehensive introductin for social scientists.* Cambridge, UK: Cambridge University Press.

Gottman, J. M. & Roy, A. K. (1990). *Sequential analysis: A guide for behavioural researchers*. Cambridge, UK: Cambridge University Press.

Greenacre, M. J. (1984). *Theory and applications of correspondence analysis.* London: Academic Press.

Griffin, W. A. & Gardner, W. (1989). Analysis of behavioral durations in observational studies of social interaction. *Psychological Bulletin*, *106*, 497-502.

Grizzle, J. E., Starmer, C. F., & Koch, G. G. (1969). Analysis of categorical data by linear models. *Biometrics*, *25*, 489-504.

Haberman, S. J. (1979). *Analysis of qualitative data, Vol. 2.* New York: Academic Press.

Harinck, F. J. H. & Hellendoorn, J. (1987). *Therapeutisch spel: Proces en interactie.* Lisse, The Netherlands: Swets & Zeitlinger.

Hout, M. (1983). *Mobility tables.* Beverly Hills: Sage.

Iacobucci, D. &Wasserman, S. (1987). Dyadic social interactions. *Psychological Bulletin*, *102*, 293-306.

Iacobucci, D. & Wasserman, S. (1988). A general framework for the statistical analysis of sequential dyadic interaction data. *Psychological Bulletin*, *103*, 379-390.

Kroonenberg, P. M., Hellendoorn, J., & Harinck, F. J. H. (1988). *Style of play in imagenary interaction play therapy: A micro analysis of sequential dyadic interaction* (Tech. Rep.). Department of Education, Leiden University.

Kroonenberg, P. M. & Verbeek, A. (1990, June). *Modelling parallel streams of observations with multinomial response models*. Paper presented at the 5th International workshop on Statistical Modelling, Toulouse, France.

Long, J. S. (1997). *Regression models for categorical and limited dependent variables.* Thousand Oaks, CA: Sage.

McCullagh, P. (1980). Regression models for ordinal data (with discussion). *Journal of the Royal Statistical Society, B*, *42*, 109-142.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers of econometrics* (pp. 105-142). New York: Academic Press.

Mood, A. M., Graybill, F. A., & Boes, D. C. (1974). *Introduction to the theory of statistics.* New York: McGraw-Hill.

Powers, D. A. & Xie, Y. (2000). *Statistical methods for categorical data analysis*. San Diego: Academic Press.

Raftery, A. E. (1995). Bayesian model selection in social research. In P. V. Marsden (Ed.), *Sociological methodology, 1995* (pp. 111-164). Cambridge, UK: Blackwells.

Rudas, T. (1998). *Odds ratios in the analysis of contingency tables.* Thousand Oaks, CA: Sage.

Theil, H. (1969). A multinomial extension of the linear logit model. *International Economic Review*, *10*, 251-259.

Theil, H. (1970). On the estimation of relationships involving qualitative variables. *American Journal of Sociology*, *76*, 103-154.

Van den Boom, D. C. (1988). *Neonatal irritability and the development of attachment: Observation and intervention.* Unpublished doctoral dissertation, Leiden University.

Vermunt, J. K. (1996). *Log-linear event history analysis: A general approach with missing data, latent variables and unobserved heterogeneity.* Tilburg: Tilburg University Press.

Vermunt, J. K. (1997a). *Log-linear models for event histories.* Thousand Oaks, CA: Sage.

Vermunt, J. K. (1997b). *ℓEM: A general program for the analysis of categorical data* (Tech. Rep.). Tilburg University.

Voeten, M. J. M. (1985). *Sequential analysis of teacher-student interaction.* Unpublished doctoral dissertation, Catholic University of Nijmegen.

Von Eye, A., Brandstädter, J., & Rovine, M. J. (1993). Models for prediction analysis. *Journal of Mathematical Sociology*, *18*, 65-80.

Von Eye, A., Brandstädter, J., & Rovine, M. J. (1998). Models for prediction analysis in longitudinal research. *Journal of Mathematical Sociology*, *22*, 355-371.

Wald, A. (1943). Tests of statistical hypotheses concerning several parameters where the number of observations is large. *Transactions of the American Mathematical Society*, *54*, 426–482.

Whittaker, J. (1990). *Graphical models in applied mathematical multivariate statistics.* Chichester: John Wiley & Sons.

Wickens, T. D. (1989). *Multiway contingency tables analysis for the social sciences.* Hillsdale, NJ: Lawrence Erlbaum.

Accepted April, 2003.

Appendix
Example of an Input File for ℓEM

Here we will discus an input file of *ℓEM*. This is the input file for the selected Model 5 for actor-specific event sampling. The input file requires a number of commands, like the number of manifest variables, the number of categories for each variable, labels for the variables and the model to be estimated. The commands are given by 3-character 'words', like `man` for the number of manifest variables. Comments can be included after a *.

*input file for the analysis of Hellendoorn and Harincks data
*$A = t_{-1}$, $B = c_{-1}$, $C = t_0$, $D = c_0$

```
man 4
* There are 4 manifest variables.
dim 4 4 4 4
* The number of categories for the first till fourth variable is 4.
lab A B C D
* The variables are named A B C and D.
mod CD|AB {CD AC AD BC cov(BD,5) wei(ABCD)}
* This line specifies the model. We have a multinomial response model with
* variables C and D as dependent and A and B as independent variables. The
* effects included are CD (included in all models), AC, AD, BC and BD for
* which we have to specify a design matrix since the weights are on this level.
* The command cov specifies that a design matrix is given to estimate the
* given interaction, together with the number of rows in the design matrix, 5.
* The design matrix follows after the command des
des
[0 1 -1 0 0 0 0 0 0 0 0 0 0 -1 1 0 *BD
0 0 0 0 1 0 0 -1 -1 0 0 1 0 0 0 0
0 0 0 0 1 0 -1 0 0 0 0 0 -1 0 1 0
0 0 0 0 0 0 0 0 -1 1 0 0 1 -1 0 0
0 0 -1 1 0 0 1 -1 0 0 0 0 0 0 0 0 ]
sta wei(ABCD)
```

* Here the weights are specified. Ones for the observed cells, zeros for the
* structural zeros.

```
[0 1 1 1 0 1 1 1 0 1 1 1 0 1 1 1
1 0 1 1 1 0 1 1 1 0 1 1 1 0 1 1
1 1 0 1 1 1 0 1 1 1 0 1 1 1 0 1
1 1 1 0 1 1 1 0 1 1 1 0 1 1 1 0
0 1 1 1 0 1 1 1 0 1 1 1 0 1 1 1
1 0 1 1 1 0 1 1 1 0 1 1 1 0 1 1
1 1 0 1 1 1 0 1 1 1 0 1 1 1 0 1
1 1 1 0 1 1 1 0 1 1 1 0 1 1 1 0
0 1 1 1 0 1 1 1 0 1 1 1 0 1 1 1
1 0 1 1 1 0 1 1 1 0 1 1 1 0 1 1
1 1 0 1 1 1 0 1 1 1 0 1 1 1 0 1
1 1 1 0 1 1 1 0 1 1 1 0 1 1 1 0
0 1 1 1 0 1 1 1 0 1 1 1 0 1 1 1
1 0 1 1 1 0 1 1 1 0 1 1 1 0 1 1
1 1 0 1 1 1 0 1 1 1 0 1 1 1 0 1
1 1 1 0 1 1 1 0 1 1 1 0 1 1 1 0]
```
* This line specifies that we use category 4 of variables C and D as baseline
* categories
```
dum -1 -1 4 4
```
* Here follow the data.
```
dat
[0 1019 338 154 0 407 53 14 0 48 83 4 0 91 39 94
877 0 591 475 406 0 201 61 56 0 146 7 152 0 93 262
332 532 0 111 68 147 0 9 70 108 0 9 38 107 0 69
149 357 89 0 34 92 11 0 6 8 12 0 227 357 112 0
0 348 58 21 0 687 74 34 0 41 39 2 0 69 6 38
361 0 212 148 643 0 214 90 39 0 105 5 84 0 57 100
47 147 0 13 77 162 0 12 24 55 0 2 18 25 0 14
8 48 5 0 23 64 5 0 1 1 3 0 15 46 11 0
0 58 57 4 0 60 29 1 0 52 89 2 0 7 12 11
55 0 98 10 30 0 40 3 38 0 147 4 13 0 15 14
98 160 0 19 35 118 0 4 106 168 0 14 19 46 0 28
4 6 5 0 3 2 4 0 3 2 6 0 15 9 8 0
0 182 31 224 0 81 11 22 0 9 20 7 0 201 78 605
149 0 99 298 73 0 19 25 10 0 28 10 209 0 167 484
56 116 0 108 18 47 0 5 10 21 0 3 104 156 0 224
160 270 71 0 39 87 10 0 11 6 24 0 627 569 249 0 ]
```