Universiteit Utrecht

Applying Hidden Markov Models in Social Sciences

S.V. Ekhart
Studentnumber: 6244386
Bachelor Sociology, Utrecht University
Bachelor's Project in Methodology and Statistics

First supervisor: Dr. E. Aarts
Second supervisor: S. Mildiner Moraga MSc

August 25, 2020

Abstract

The increasing availability of intense longitudinal data in social sciences asks for appropriate analysis methods. This thesis focuses on the use of Hidden Markov Models in social and behavioural sciences. Hidden Markov Models are not widely known/used, which may have to do with unfamiliarity or perceived limitations under researchers. This thesis will aim to give some insights on what can be gained from using a Hidden Markov Model. The analysis in this thesis is done with data about children's behaviour towards other children or robots. A covariate (if the child is interacting with another child or a robot) is taken into account. A Hidden Markov Model and a chi-squared test were both applied on the data and the outcomes were compared. Both the Hidden Markov Model and the chi-squared test revealed differences in behaviour between children interacting with other children or children interacting with a robot. The Hidden Markov Model gave extra information about children's transitions between different types of behaviour, compared to the chi-squared test. There were also differences found between the two conditions on specific variables. The heterogeneity which is allowed by the Hidden Markov Model also proved to have added value. A few assumptions in preparation of the data were made. This could influence the outcomes of both the Hidden Markov Model and the chi-squared test. For further research it would be interesting to add more covariates in the analysis to reduce the unexplained variance. It would also be interesting to look at opportunities to make Hidden Markov Models more accessible for researchers.

Preface

The preface of a bachelor thesis is probably not the part that catches the most interest. Unfortunately, because this specific part provides a unique and valuable opportunity for the reader to get to know the author and his specific perspective on the subject. In this case I want to give you, the reader of my Bachelor thesis, some information about my personal and professional history.

About two years ago I started with the Bachelor Sociology after a year of mathematics. During the last years I discovered my interest was not in mathematics only, but in the combination of abstract models with mathematics behind them and social behaviour. Explaining the complexity of social behaviour in the world around us is one of the things I keep admiring. To get more knowledge about this type of research I chose the subjects with the combination of modelling and social behaviour. Also, I did a minor in complex systems studies which sparked my interest in different methods (including less conventional) that can be applied on social data.

In January I needed to choose which subjects I would like to work on during my thesis. Again, I saw that from all subjects my interest was in the more abstract field of social sciences. Therefore, I started to work on a Bachelor thesis about the use of Hidden Markov Models in social sciences.

I want to thank Emmeke and Sebastian very much for their adequate help and inspiration during the process. Also during these turbulent times related to corona they supported me in the process. Furthermore, I want to thank my friends, parents, step-father and partner for their help en support over the last few months. In the quarantine period their positive moral support and stories about their own thesis (over the phone) helped me to stay motivated.

Suzanne Veerle Ekhart
Utrecht, August 25, 2020

Contents

1 Introduction

Hidden Markov Models (HMMs) can be found in a wide variety of fields such as in bioinformatics (Eddy, 1998), earth-science (Beyreuther, Hammer, Wassermann, Ohrnberger, & Megies, 2012), economics (Yu & Sheblé, 2006) and also in social sciences such as psychology (Visser, Raijmakers, & Molenaar, 2002) and pedagogy (Fok, Wong, & Chen, 2005). HMMs are models used to analyse intense longitudinal data. This is a type of time series data with a large number of data points for each individual on each variable (in section 1.1 there will be a more detailed description of this type of data). Lately, due to increasing amount of available intense longitudinal data and improvements in computational technology, HMMs have been receiving more and more attention as an interesting type of modelling during research (Schafer, 2006). Still, in psychology Markov Models are not often used, which may have to do with unfamiliarity with the models or the perceived limitations/challenges in the use of the models under psychology researchers (de Haan-Rietdijk et al., 2017).

HMMs can be useful when analysing time series data from individual actors, such as persons in social sciences, or DNA sequences in bioinformatics. The HMM gives the opportunity to analyze this data from individuals without relying on cross-sectional data. HMMs are also very flexible compared to other time series models. (Visser, 2011)

An example of a research area in social sciences where an HMM is already used is the research about alcoholism treatment by Sherley et. al. (2010). The aim of this research was to describe the drinking behaviour of the subjects using (Hidden) Markov Models. The use of an HMM makes it possible to describe processes that contain sudden changes. Many statistical models make the assumption that behaviour is stationary, which is not applicable in the case of alcoholism treatment data where the behavioural patterns contain sudden changes most of the time. (Shirley et al., 2010)

The result of using an HMM in the research mentioned above about alcoholism treatment was a new definition of relapse, based on the observable mental and physical state of the subject instead of the direct observable drinking pattern. This new definition is also conceptually supported by the cognitive-behavioural model of relapse. For clinical implementation this new way of looking at relapse could be compared to already existing definitions and in this way make alcoholism treatment more efficient. (Shirley et al., 2010)

Another example of HMM use in social and behavioural science is the research from Hao-Chuan Wang (2008) about idea generation. In the research the HMM proveded a new way to find out more about the hidden structures and processes in brainstorming groups and idea generation. Also, with the use of HMMs they were able to look into the influence of social interactions on the dynamics of brainstorming groups. (Wang, 2008)

An example in the behavioural science about animal behaviour can be found in the research on at-sea behaviour of a pelagic seabird. The HMM offered insights about the behaviour without categorization the data in discrete groups of observations prior to the data analysis. Using the HMM the researchers found differences in foraging trip behaviour which may have consequences for the breeding of the birds. This information can, for example, have added value for conservation planning by identification of important areas and habitats for the animals. (Dean et al., 2013)

The increasing availability of intense longitudinal data in social and behavioural science asks for an appropriate statistical method. In this thesis I will look at what can be gained from using an HMM in research with intense longitudinal data compared to the more conventional statistical methods. In this thesis I will compare the outcomes of an HMM with the outcomes of a chi-squared test. To do so I will first explain the principle of intense longitudinal data and I will introduce the (multilevel) Hidden Markov Model and the chi-squared test.

1.1 Intense longitudinal data

A longitudinal data set consists of repeated observations for one or more variables for different subjects (Zeger & Liang, 1986). An example of a longitudinal data set is a data set where there are multiple observation moments for the subjects such as panel surveys, clinical trials or studies about human development. In these types of research just a few measurements suffices. (Schafer, 2006).

Intense longitudinal data is a type of time series data with many observations over time. Intense longitudinal data gives us the opportunity to learn about the dynamics of a process and not just the starting and end point. (Hamaker & Wichers, 2017)

1.2 Research question

The goal of this thesis is to compare the insights researchers can get using a Hidden Markov Model compared to summary statistics using the PInSoRo data set as a case study. In summary statistics, the dynamics of a system is hard to study. It is hard to examine how much the values of the variables of individuals are changing over time because the analysis is often done over the summaries of the time series instead of the whole series. HMMs allow modelling insights about the dynamics of the time series data, it is a quantitative research method with theoretical interpretable outcomes (Shirley et al., 2010). These theoretical interpretations can be useful in real life situations such as (in the context of the data) the use of robots in different situations with diverse aims.

The research question that will be answered is: *What additional insights can be gained from using a multilevel Hidden Markov Model output compared to static statistical tests?*

1.3 Context of the data

In a world with an emerging amount of robots in many different fields, questions about human interaction with these robots arise. These questions can be about what can be done to improve interactions between humans and robots (Belpaeme, Baxter, Read, et al., 2013) and which difficulties can arise in the process of improvement (Belpaeme, Baxter, De Greeff, et al., 2013)(Dautenhahn, 2007). In this thesis the interactions between children and robots will be featured. Looking at children in combination with robots can be interesting because children are likely to see life-like character traits which possibly makes the interaction more natural. Robots can be useful in many different areas such as healthcare and education. In healthcare, animals are sometimes used to improve the well-being of children in a hospital. But since this method is not always sustainable, robots might be an interesting alternative. (Belpaeme, Baxter, De Greeff, et al., 2013)

The data used in this thesis is about children who interact with other children or robots. The data is further explained in section 2.

1.4 The Hidden Markov Model

Now the principle of HMM will be clarified and there will be given an (simplified) example. In an HMM an actor can be in a certain state and this state can be described by (multiple) observations. This state can change over time and the dependence of the data over time is thus an important part of the HMM. This dependence (shown in figure 1) is as follows: the state S_t is dependent on the previous state S_{t-1} but not on the other states, which is called the Markov property. At each state we have a real life observation (Y_t).(Visser, 2011)

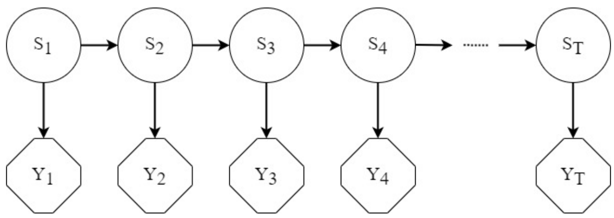


Figure 1: Dependency graph for the HMM. (Based on Visser (2011))

For an HMM it is interesting to know what the probabilities are for moving from one state to another. The transition probability matrix $\mathbb{A}$ shows this probability, where each row contains probabilities of moving to another state. a_{xy} contains the probability of moving from state x to state y. Each row needs to sum to one ($\sum_{y=1,...,n} a_{xy} = 1$). (Visser, 2011)

$$\mathbb{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1y} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2y} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3y} \\ \vdots & \vdots & \vdots & . & \vdots \\ a_{x1} & a_{x2} & a_{x3} & \cdots & a_{xy} \end{pmatrix}$$

When there is a deterministic one-to-one function, which couples values of S_t with values of Y_t, there is no case of an HMM because the states are not hidden (then we would not want to use a *hidden* Markov model but an *observed* Markov model). In practace, values S_t cannot be directly observed from the values of Y_t quite often. When this is not possible, but a probabilistic relationship between states and observations can be inferred, an HMM can be used. This probabilistic relationship can be shown in a conditional observation probability matrix $\mathbb{B}$, which gives the probabilities which observation $(Y_1, Y_2, ..., Y_T)$ will be observed in a certain state $(S_1, S_2, ..., S_T)$. This matrix is shown below where p is the probability of observing a certain observation in a certain state. (Visser et al., 2002)(Visser, 2011)

$$\mathbb{B} = \begin{pmatrix} p_{S_1Y_1} & p_{S_1Y_2} & \cdots & p_{S_1Y_T} \\ p_{S_2Y_1} & p_{S_2Y_2} & \cdots & p_{S_2Y_T} \\ \vdots & \vdots & . & \vdots \\ p_{S_TY_1} & p_{S_TY_2} & \cdots & p_{S_TY_T} \end{pmatrix}$$

When using an HMM it is required to determine the starting values for the MCMC (the algorithm that is used). This is done by the researcher in the form of a probability distribution over the possible states. In this thesis I used a discrete univariate probability distribution with finite support for the distribution of the emissions. This distribution results in a distribution with the same probability for observing all the possible observations, where the probability of observing (C) a certain value is determined by the formula shown in equation 1. (Visser, 2011)

$$C = \frac{1}{q} \tag{1}$$

With q = number of possible observations for variable

The starting values of the transition probability matrix in this thesis is a matrix with the value 0.8 on the diagonals. This results in a relatively low probability of changing from one state to another.

Also, the number of states used in an HMM needs to be determined. This choice is based on the AIC (Akaike Information Criterion) which provides a means for the selection of a certain model. This selection is based both on how complicated the model is, and how well it fits the data. Besides the AIC, there also needs to be looked at the content of the states a model produces and how reasonable and theoretically interpretable the results are.(Aarts, 2019b)

Now I will clarify HMMs using a fictional example. We look at a situation from two fictional friends. The two friends are living far away from each other and they talk to each other over the phone. Friend A does one of 3 activities (walking (Y_1), shopping (Y_2) and cleaning (Y_3)) depending on the weather. Friend B does not have any information about the weather except from general trends. Based on the activities that friend A does friend B tries to guess the weather from that day. The hidden states in the weather could be sunny ($_1$) or rainy (S_2), but it is not possible to observe them directly, they are hidden. In this example the states represent the weather (S) and the activities are observations (Y).(Mackenzie, 2010)

Now we can compute the transition probability matrix $\mathbb{A}$ with the probabilities of moving from one state to another (for this example there are given fictional probabilities). In the case of this example this is a matrix with the probability of moving from the rainy to the sunny state:

$$\mathbb{A} = \begin{pmatrix} 0.75 & 0.25 \\ 0.30 & 0.70 \end{pmatrix}$$

Now we determine if the relationship between the observations and the hidden states is probabilistic. In the example it is because there is not a deterministic one-to-one function. After determining if there is a probabilistic relationship, we can make a conditional observation probability matrix $\mathbb{B}$ which gives the probabilities with which the observations will be observed in a certain states (for this example there are given fictional probabilities):

$$\mathbb{B} = \begin{pmatrix} 0.45 & 0.40 & 0.15 \\ 0.10 & 0.35 & 0.55 \end{pmatrix}$$

In the fictional example from the two friends, this would suggest that when observing walking, there is a high probability that the weather is in state 1 (sunny). When observing cleaning, the fictional results suggest that there is a high probability that the weather is in state 2 (rainy).

1.5 Multilevel Hidden Markov Model

In this project there will be looked into multilevel Hidden Markov Models. When a data set consists of observations collected from multiple subjects, two different levels can be used to fit an HMM.

The first method is looking at an HMM for each subjects. This method is very computationally intensive and the interpretation is hard due to the large number of parameter estimates. The second method is performing an HMM for all the subjects together. This requires us to make large assumptions about the homogeneity between the subjects in the parameters, which is usually not eligible. To analyse multiple subjects, a multilevel HMM can be used. (Aarts, 2019b)

In a multilevel HMM there are two levels of parameters. The first level consists of the observations within subjects where each subject has its own parameter but they all follow the same group distribution. The parameter sets for the specific subjects are realizations of a common distribution on group level. The first level is the group distribution and the second level are the individual parameter sets. (Aarts, 2019a)

In the context of the transition probability matrix and the conditional distribution matrix this would mean that there can be heterogeneity in these matrices for the different subjects but that there is one HMM estimation. (Aarts, 2019b)

In earlier (and more conventional) research, the different levels are moved to one single level. This single level was analysed by a conventional statistical method. This is, however, inadequate, and leads to both statistical and conceptual problems. (Hox, Moerbeek, & Van de Schoot, 2018)

Also, most of the more conventional methods make an assumption of independency of the observations, which is often not the case. Violating this assumption can lead to false significant results. (Hox et al., 2018)

When using multilevel models it is possible to analyze variables at different levels simultaneously with inclusion of dependencies. This is not possible with most of the conventional statistical methods and thus a reason to do multilevel research when dealing with data on different levels. (Hox et al., 2018)

1.6 Conventional static statistical tests

Statistics is a very important part of research in social sciences. Most researchers use statistical tests they already know. In this thesis the HMM will be looked into as an informative alternative for these static statistical test in research with intense longitudinal data.

The Hidden Markov Model will be compared with a chi-squared test. This test can be used because the type of data used in this project meets the condition ”the level of measurement of all the variables is nominal or ordinal”. There are some assumptions that need to be made when performing a chi-squared test. First, the data needs to be given in frequencies and not in percentages (or another transformation of the data). Second, each subject has just one value for all the variables. Third, a chi-squared test can not be used when subjects are compared over time. Fourth, the groups which will be compared must be independent. Fifth, the two variables are categories (norminal or ordinal). Last, the expected values should be above five in more than 80 percent of the cells. Also, none of the cells should have an expected value of less than one. (McHugh, 2013)

When determining the fit of a linear model the squared differences between the observed values and the predicted values can be summed (see equation 2).(Field, 2017)

$$\text{Total error} = \sum_{i=1}^{n} (observed_i - model_i)^2 \tag{2}$$

For the chi-squared test (with categories instead of a linear model) a similar method is used. But now the deviation for each of the observations is standardized by dividing the sum of squares by the degrees of freedom (see equation 3). (Field, 2017)

$$\chi^2 = \sum \frac{(observed_{ij} - model_{ij})^2}{model_{ij}} \tag{3}$$

The rest of this thesis is structured as follows. Section 2 provides information about the methods used and information about the preparation of the data. Section 3 first gives an overview of summary statistics and outcomes of the chi-squared test and the outcomes of the Hidden Markov Model. Section 4 will present the conclusions in terms of answering the research question and will discuss the limitations of this project. This section also gives some suggestions for further research.

2 Methods

The first stage of the project consisted of a data-set search to derive a suitable data-set. This search was hard because of the confidentiality (due to privacy reasons) of potential data. The Plymouth Interactive Social Robots (PInSoRo) data-set (Lemaignan, Edmunds, Senft, & Belpaeme, 2018) was eventually used during this project.

2.1 Data description

The data was collected in a situation where two children or a child and a robot were sitting in front of each other (face-to-face) with a large, horizontal interactive touchscreen between them. The pairs were placed at the table and they could freely draw and interact with the displayed objects on the touchscreen (see figure 2). The time for free playing was 40 minutes at maximum. The exact data acquisition protocol is added in appendix A.1.

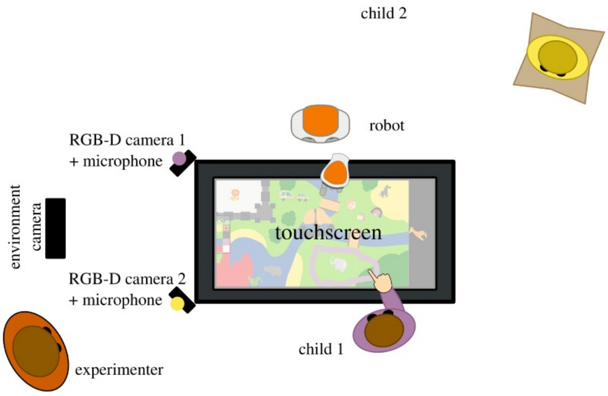


Figure 2: Experimental setup used during data collection (Lemaignan et al., 2018)

During the interactions of the pairs, information was gathered by a camera and microphone. After the data collection, five annotators (with one hour of training) annotated the data. The data was coded for three different dimensions of social dynamics (task engagement, social attitude and social engagement).

Purple child task engagement contained four labels: goal oriented, aimless, adult seeking and no play. *Purple child social attitude* contained five labels: pro-social, adversarial, assertive, frustrated, passive/bored. *Purple child social engagement* contained also five labels: solitary, onlooker, parallel, associative and cooperative.

The raw data consisted of cells with more than one value because some pairs were annotated by more than one annotator. For the analysis, I simply used the first value because there wasn't found a logical order (such as alphabetical) in the sequence of the values. It seemed the most reasonable that if there were more values, but also overlapping values, the value which was given most was written down first in the cell. This would support the decision of taking the first, but there wasn't real support for this in the article added to the original data (Lemaignan et al., 2018).

In addition, I applied a rolling-based filtering of the data such that small changes in behaviour for a small time were filtered out and transitions in behaviour were still in the data. This was done by checking for every time-step the time-step itself and the four following time-steps. The value that was present the most was given to the time-step. After that, for each subject the last four time-steps were deleted.

Finally, the data of subjects which ended with a long time of missing data at all variables were cut off after the last non missing data time-step at any of the variables.

In figure 3, one of the data subjects is shown in an interpretable plot after changing the data. There can be seen data from a child who is interacting with another child where each color represents a different value of the variables. When performing an HMM it is interesting to look at which combinations of values are in a state (i.e. the content of the states).

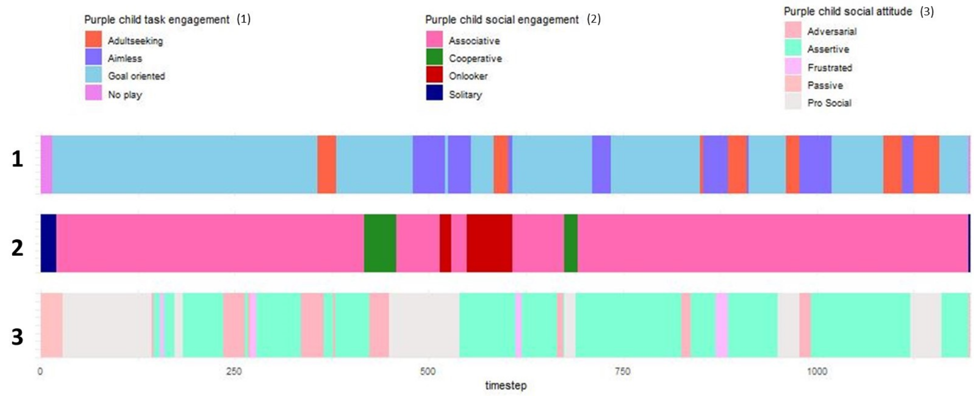


Figure 3: Plot of the observations from id number 2 (child-child).

2.2 Data analysis

During the data analysis I performed a chi-squared test on the variables task engagement, social engagement and social attitude in combination with the variable condition (child-child or child-robot). I obtained the frequencies by counting the number of times a value was measured over the whole data file. This results in group level frequencies and not apart for all subjects.

I fitted the HMMs with two, three, four and five hidden states with the variable condition as a covariate for both the transition probabilities and the condition probabilities. I used the statistical software R and the package mHMMbayes to fit a multilevel HMM in a Bayesian framework. Also, there will be expanding R-code to transform and visualize the outcomes of the HMM. A part of this code is written especially for this project, other parts are written also for further research.

The transition probabilities were calculated using the package mHMMbayes. In this package the subject specific probability γ_{kij} are calculated using a linear predictor function with $m-1$ (with m equals the number of states) random intercepts to allow the earlier said heterogeneity in the outcomes. The transition probabilities were modeled with equation 5 with transitions from state $i \in 1, 2, \ldots m$ to state $j \in 1, 2, \ldots, m$ which are modeled using m batches of $m-1$ random intercepts (see equation 4). By extending the subject-specific state-dependent probabilities to a multinomial logit (MNL) model, inclusion of random effects and covariates is allowed. (Aarts, 2019a)

The subject specific probability consists of a group level mean ($\bar{a}_{(S)ij}$) and an error term ($\psi_{(S)kij}$) (see equation 6) (Lynch, 2007). This group level mean is determined by a normal distribution with hyperpriors. The distribution for the covariance follows an Inverse Wishart distribution (Aarts, 2019a).

$$\boldsymbol{\alpha}_{(S)ki} = (\alpha_{(S)k13}, \ldots, \alpha_{(S)k1m}, \alpha_{(S)k23}, \ldots, \alpha_{(S)k2m}, \ldots, \alpha_{(S)km2}, \ldots, \alpha_{(S)km(m-1)}) \\ \text{With } k = \text{number of subjects} \tag{4}$$

$$\gamma_{kij} = \frac{\exp(\alpha_{(S)kij})}{1 + \sum_{\bar{j} \in Z} \exp(\alpha_{(S)ki\bar{j}})} = \text{MNL}(\alpha_{(S)kij}) \tag{5}$$

$$\alpha_{(S)kij} = \bar{\alpha}_{(S)ij} + \psi_{(S)kij} \tag{6}$$

The probability of observing categorical outcome l within state i on a subject specific level (θ_{kil}) was calculated using a linear predictor function. This function (see equation 7) consists of $q-1$ random intercepts (each categorical outcome q except for the first one, for model identification reasons, has its own intercept). For the conditional probability matrix, the heterogeneity in the model was accommodated by random intercepts. The subject specific probabilities of observing categorical outcome $l \in 1, 2, \dots, q$ in state $i \in 1, 2, \dots, m$ is modeled using m batches of $q-1$ random intercepts (see equation 8). (Aarts, 2019a)

Again, the subject specific probability consists of a group level mean ($\bar{a}_{(O)ij}$) and an error term ($\psi_{(O)kij}$) (see equation 9) (Lynch, 2007). This group level mean is determined by a normal distribution with hyperpriors. The distribution for the covariance follows an Inverse Wishart distribution (Aarts, 2019a).

$$\begin{gathered}\boldsymbol{\alpha}_{(O)ki} = (\alpha_{(O)ki2}, \alpha_{(O)ki3}, \dots, \alpha_{(O)kiq}) \\ \text{With } k = \text{number of subjects}\end{gathered} \tag{7}$$

$$\theta_{kil} = \frac{\exp(\alpha_{(O)kil})}{1 + \sum_{\bar{l}=2}^{q} \exp(\alpha_{(O)ki\bar{l}})} = \mathrm{MNL}(\alpha_{(O)kil}) \tag{8}$$

$$\alpha_{(O)kij} = \bar{\alpha}_{(O)ij} + \psi_{(O)kij} \tag{9}$$

The significance of both the transition and conditional probabilities was tested using the 95% credibility intervals obtained from the MCMC samples to test the statistical significance of the parameters of the model.

3 Results

In this section the results of the chi-squared test and the HMM will be shown. First the summary statistics will be presented, then the outcomes of the chi-squared test and lastly the results of the HMM analysis.

3.1 Summary statistics

After cleaning the data I computed the proportions from the different annotations with a split between the two different conditions.

Looking at the proportion in the variable *purple child social attitude* in figure 4, we see that in the child-child condition both pro social and passive are the most frequent annotations; this behaviour is highly presented in the data. For the child-robot condition pro social and assertive behaviour are less frequent, whereas passive behaviour is now a more important component of the behaviour. Frustrated behaviour seldom occurs in the child-child condition. The differences between the two groups are confirmed by the chi-squared test ($\chi^2(5) = 18279$, $p < 2.2e^{-16}$).

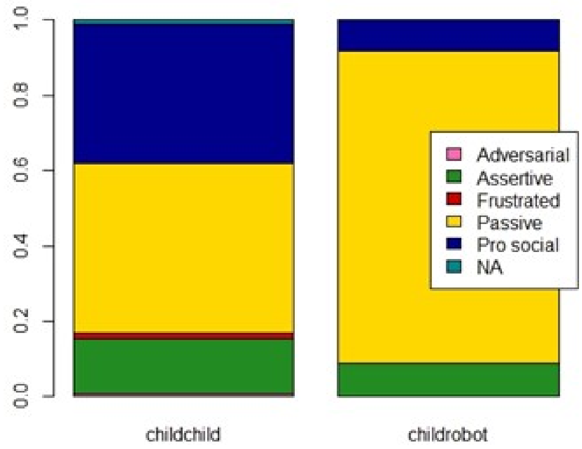


Figure 4: *Purple child social attitude* by condition

In the variable *purple child social engagement* there can be seen that the annotators annotated cooperative behaviour in the child-child condition but not in the child-robot condition (exact tables can be found in appendix A.3). The chi-squared test confirms the differences between the two conditions and supports the idea that social engagement is dependent on condition ($\chi^2(5) = 10761$, $p < 2.2e^{-16}$).

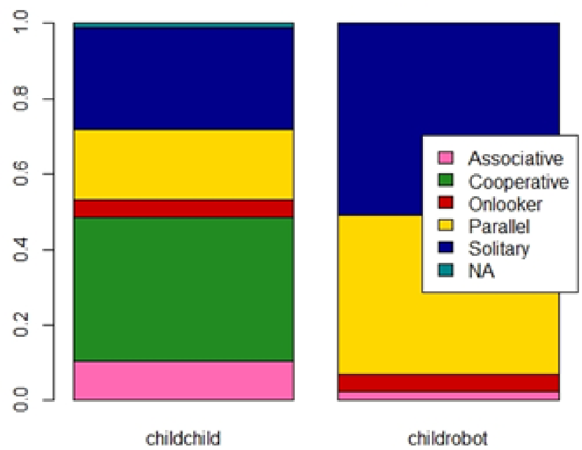


Figure 5: *Purple child social engagement* by condition

Looking at the variable *purple child task engagement* there can be seen that the proportions are for both conditions about the same. For the child-child condition there is more no play and less aimless and goal oriented annotated. The outcomes from the chi-squared test also suggest that there is dependency from task engagement on condition ($\chi^2(4) = 2122.8$, $p < 2.2e^{-16}$).

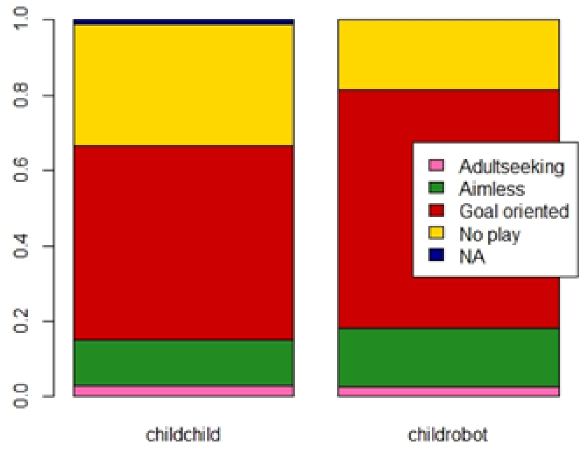
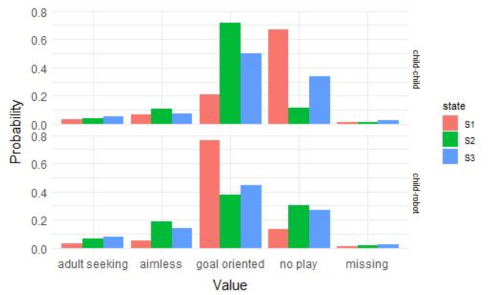


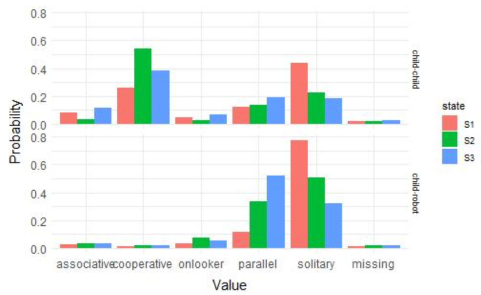
Figure 6: *Purple child task engagement* by condition

3.2 Outcomes HMM

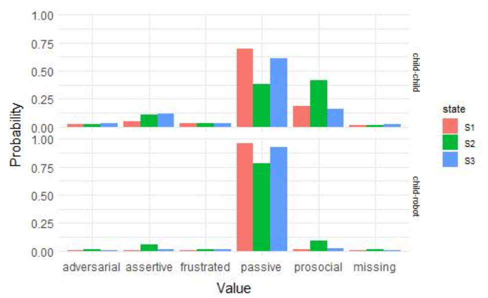
The models with a different number of hidden states m produced different AIC's. Even though the five state model gives the lowest AIC, I decided to opt for a simpler model with less states which has an improved interpretability compared to the other models. Figure 7 shows the different in the content of the states and the overall model is also not too complicated. The outcomes of the two, four and five state model are added in the appendix.



(a) Conditional probabilities *purple child task engagement* by condition



(b) Conditional probabilities *purple child social engagement* by condition



(c) Conditional probabilities *purple child social attitude* by condition

	S1toS1	S1toS2	S1toS3
child child	0.9609	0.0212	0.0179
child robot	0.9705	0.0144	0.0152
	S2toS1	**S2toS2**	**S2toS3**
child child	0.0181	0.9645	0.0174
child robot	0.0126	0.9715	0.0156
	S3toS1	**S3toS2**	**S3toS3**
child child	0.0156	0.0186	0.9655
child robot	0.0208	0.0207	0.9578

(d) Transition probabilities for an HMM with three states

Figure 7: Output from the Hidden Markov Model with three states ($AIC = 2560.626$).

Looking at the variable *purple child task engagement* for state one, no play has the highest probability for the child-child conditions. On the contrary, for the child-robot in state one observing goal oriented behaviour has the highest probability (see figure 7a). In the second state observing goal oriented behaviour has the highest probability in both conditions, but in the child-robot condition also no play has a similar probability. In the third state for both conditions the highest probability is to observe goal oriented behaviour, but no play behaviour has a high probability as well.

For the variable *purple child social engagement* we see differences between the two conditions (see figure 7b). For the child-robot condition the probability of observing cooperative behaviour is very low. On the contrary, for the child-child condition this type of behaviour has in all states a relatively high probability. In state one, for the child-child condition, there is for both cooperative behaviour and solitary behaviour a relatively high probability. For the child-robot condition we see that solitary behaviour has the highest probability. In state two we see for the child-child condition a high probability of observing cooperative behaviour. For the child-robot condition in state two there is a similar probability to observe parallel or solitary behaviour. In state three there can be seen that, for the child-child condition, there is a high probability of observing cooperative behaviour, but also parallel and solitary. For the child-robot condition it is just a mix of parallel and solitary social engagement.

The last variable, *purple child social attitude* (see figure 7b) is almost the same for both conditions. For state one, two and three there is a high probability of observing a passive attitude. For the child-robot condition however there is also in all states a noteworthy probability to observe pro social behaviour. Especially in state two this probability is very high (higher then the probability of observing passive behaviour).

Summarizing this we see the following states. The states in the child-child condition are shown on the left and the states in the child-robot condition are shown on the right.

Child-child

State 1 Child does not play, has a high probability to be solitary but sometimes is cooperative and the child shows passive behaviour towards the other child.

State 2 Child is goal oriented, cooperative and shows varying passive or pro social towards the other child.

State 3 Child is goal oriented, cooperative and passive towards the other child.

Child-robot

State 1 The child is goal oriented but plays alone and is passive towards the robot.

State 2 The child can be goal oriented or is not playing it does it alone or parallel towards the robot and it shows passive behaviour towards the robot.

State 3 The child is mostly goal oriented and it mostly plays parallel towards the robot but sometimes also completely solitary, the child shows passive behaviour.

In figure 7d the transition probabilities are shown. These probabilities represent the odds of moving from one state to another. None of the differences between the different conditions have been found significant.

4 Conclusion and discussion

In this thesis the question I wanted to answer was: *What additional insights can be gained from using a multilevel Hidden Markov Model output compared to static statistical tests?* To answer this question I compared the results of a chi-squared test with the outcomes of an HMM.

Looking at the results as shown in section 3.1 and 3.2 we can see that there are differences in behaviour between the two conditions. Differences can be observed from the results from both the HMM and the chi-squared test

Looking at the outcomes of the HMM we can conclude that the transition probabilities for the different conditions are not significantly different. This means that we can not conclude that the transition probabilities (so the change between different states with different content) differ between the two conditions (child-robot and child-child). The high probabilities in the cells with probability for not moving to another state means that children are having a low probability of changing their behaviour. The probability of showing the same behaviour the whole time is very high (between 95 and 98 percent). This is information that is not possible to get using conventional static statistical tests (such as the chi-squared test).

The HMM revealed differences between the conditions on specific categories from the variables. For the variable *purple child task engagement* we see differences in the content of the states between the two conditions. In state one, the probability of observing no play behaviour is very high for the child-robot condition but for the child-child condition observing goal oriented behaviour has the highest probability.

The outcomes of the HMM also revealed differences in the variable *purple child social engagement*. For the child-robot condition the probability of observing cooperative behaviour is almost zero in all states, but this probability is noteworthy in all states in the child-child condition. For state two and three in the child-robot condition the probability of observing parallel social engagement is also much higher then in the child-child condition.

For the social attitude it can be seen that there is a very small probability to observe pro social behaviour in all states for the child-robot condition. For the child-child condition all states have a relatively high probability to observe pro social behaviour or passive behaviour compared to the child-robot condition, especially state two.

The above mentioned differences are differences on specific categories; the HMM allows heterogeneity in the data. The chi-squared test also revealed significant differences between the two conditions. The differences that are found with a chi-squared test are overall changes between the two conditions so this gives less (or different) information than the outcomes of the HMM.

A limitation of this project was the small sample size. In this project there were not many significant results which can be caused by lack of statistical power. It would be interesting to perform the HMM on bigger samples to look whether significant results can be found.

It would be interesting to include additional covariates on the model. In this thesis I have looked at the covariate condition (child-robot and child-child) but it would be interesting to look at more covariates such as age or gender. This could perhaps increase the statistical power and reduce the unexplained variance.

Another limitation in this project is the assumptions that needed to be made in the preparation of data. The assumption that the first word in a cell is taken into account during the analysis for example. This can influence the results. These assumptions could be wrong and this would mean that some of the data-points are not accurate and thus the results could be inaccurate.

For further research it would also be interesting to look at the application of HMM in more fields of social sciences. It would be interesting to look for research subjects where HMMs are very applicable but not yet used. Also, it would be interesting to look at more ways of gathering intense longitudinal data in an efficient way.

Besides using HMMs in more different fields in social sciences and gathering intense longitudinal data more efficiently, it would also be interesting to look at ways to make them more accessible for researchers in all the fields in social sciences. As said earlier, due to unfamiliarity and perceived limitations and challenges in the use of these models, researchers do not often use them. The development of the mHMMbayes package in R is in my opinion a very good start to make it easier for researchers to use a Hidden Markov Model in their research. It would be desirable to refute the perceived limitations and challenges social and behavioural researchers feel regarding HMMs, and guide them when they want to use Hidden Markov Models during their research.

A Appendix

A.1 Data acquisition protocol

The data acquisition protocol adopted from (Lemaignan et al., 2018).

Greetings (about 5 min)

- Explain the purpose of the study: showing robots how children play
- Briefly present a Nao robot: the robot stands up, gives a short message (Today I'll be watching you playing in the child-child condition;Today I'll be playing with you in the child-robot condition), and sits down.
- Place children on cushions.
- Complete demographics on the tablet.
- Remind the children that they can withdraw at anytime.

Gaze tracking task (40 sec)
Children are instructed to closely watch a small picture of a rocket that moves randomly on the screen. Recorded data is used to train a eye-tracker post-hoc.

Tutorial (1-2 min)
Explain how to interact with the game, ensure the children are confident with the manipulation/drawing.

Free-play task (up to 40 min)

- Initial prompt: "Just to remind you, you can use the animals or draw. Whatever you like. If you run out of ideas, there's also an ideas box. For example, the first one is a zoo. You could draw a zoo or tell a story. When you get bored or don't want to play anymore, just let me know."
- Let children play.
- Once they wish to stop, stop recording.

Debriefing (about 2 min)

- Answer possible questions from the children.
- Give small reward (e.g. stickers) as a thank you.

A.2 Data codebook

Table 1: Codebook of the codes used in the recoded data-set

Variable	**Value**	**Value name**
Task engagement	1	Adultseeking
	2	Aimless
	3	Goal Oriented
	4	No play
	5	Missing
Social engagement	1	Associative
	2	Cooperative
	3	Onlooker
	4	Parallel
	5	Solitary
	6	Missing
Social attitude	1	Adversarial
	2	Assertive
	3	Frustrated
	4	Passive
	5	Prosocial
	6	Missing

A.3 Summary statistics

Table 2 shows the frequencies from the variable *purple child task engagement* after data selection by condition.

Table 2: *Frequency statistics purple child task engagement after data selection*

		Condition	
		child-child	child-robot
Purple child task engagement	Adultseeking	0.03003600	0.02706152
	Aimless	0.12058356	0.15527531
	Goal oriented	0.51345864	0.63337462
	No play	0.32528885	0.18428856
	Missing data	0.01063295	0.00000000

Table 3 shows the frequencies from the variable *purple child social engagement* after data selection by condition.

Table 3: *Frequency statistics purple child social engagement after data selection*

		Condition	
		child-child	child-robot
Purple child social engagement	Associative	0.10419457	0.02150659
	Cooperative	0.38012810	0.00000000
	Onlooker	0.04902043	0.04695417
	Parallel	0.18553248	0.42198701
	Solitary	0.27024029	0.50955223
	Missing data	0.01088413	0.00000000

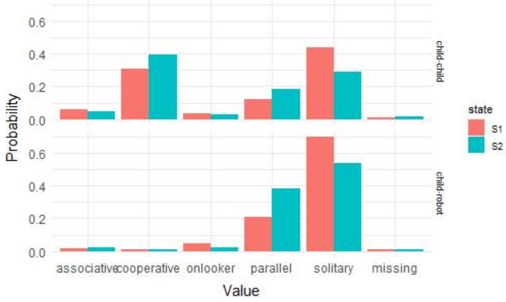
Table 4 shows the frequencies from the variable *purple child social attitude* after data selection by condition.

Table 4: *Frequency statistics purple child social attitude after data selection*

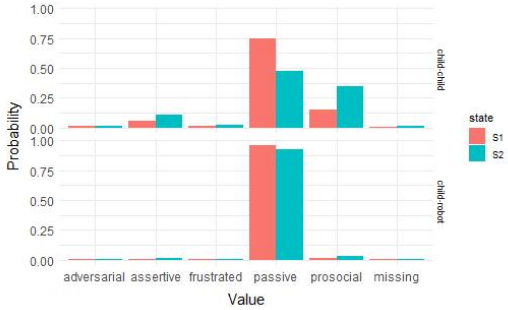
		Condition	
		child-child	child-robot
Purple child social attitude	Adversarial	0.00877009	0.00000000
	Assertive	0.14390070	0.08658935
	Frustrated	0.01463078	0.00000000
	Passive	0.45070747	0.82914837
	Pro social	0.37104404	0.08426228
	Missing data	0.01094692	0.00000000

A.4 Outcomes HMM

(a) *Purple child task engagement*



(b) *Purple child social engagement*



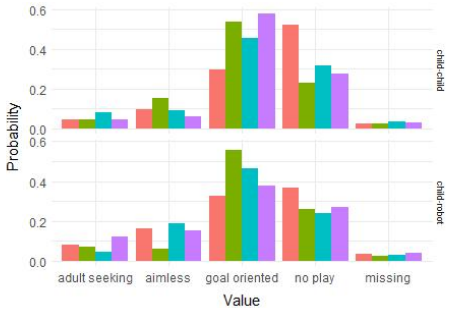
(c) *Purple child social attitude*

	S1toS1	S1toS2
child child	0.9803	0.0197
child robot	0.9803	0.0197

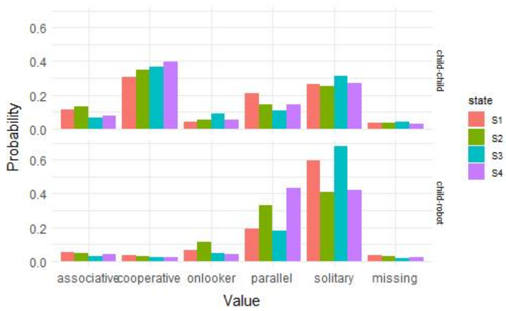
	S2toS1	S2toS2
child child	0.0176	0.9824
child robot	0.0197	0.9803

(d) Transition probabilities for an HMM with two states

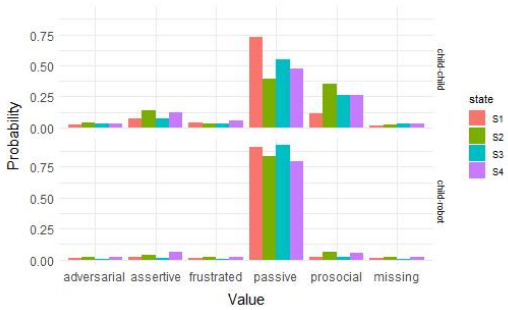
Figure 8: Condition and transition probabilities from the Hidden Markov Model output with two states ($AIC = 3236.94$).



(a) *Purple child task engagement*



(b) *Purple child social engagement*

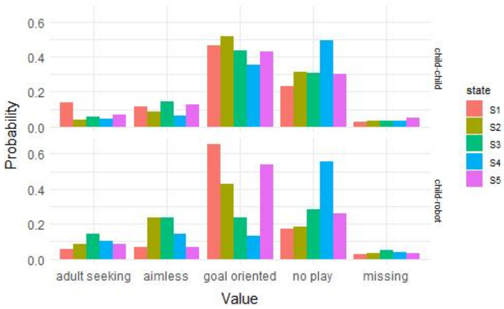


(c) *Purple child social attitude*

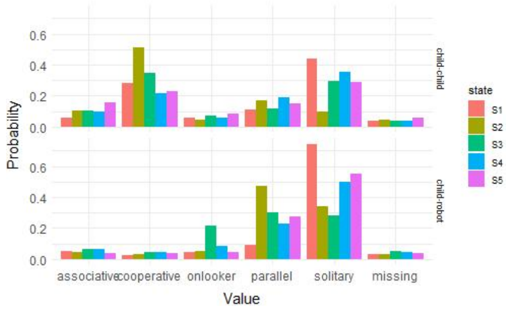
Figure 9: Condition probabilities from the Hidden Markov Model output with four states ($AIC = 2212.106$).

	1 to 1	**1 to 2**	**1 to 3**	**1 to 4**
Child-Child	0.9512	0.0198	0.0120	0.0169
Child-Robot	0.9403	0.0168	0.0234	0.0191
	2 to 1	**2 to 2**	**2 to 3**	**2 to 4**
Child-Child	0.0190	0.9521	0.0139	0.0153
Child-Robot	0.0155	0.9372	0.0267	0.0201
	3 to 1	**3 to 2**	**3 to 3**	**3 to 4**
Child-Child	0.0129	0.0159	0.9530	0.0181
Child-Robot	0.0160	0.0173	0.9504	0.0154
	4 to 1	**4 to 2**	**4 to 3**	**4 to 4**
Child-Child	0.0203	0.0186	0.0169	0.9438
Child-Robot	0.0162	0.0215	0.0188	0.9436

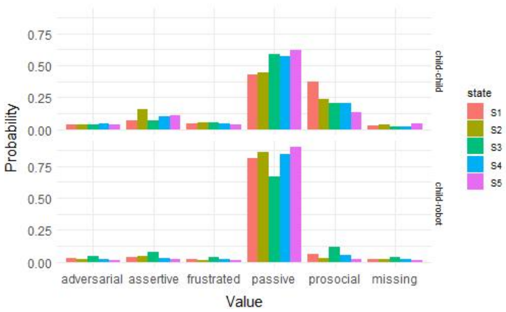
Figure 10: Transition probabilities for an HMM with four states



(a) *Purple child task engagement*



(b) *Purple child social engagement*



(c) *Purple child social attitude*

Figure 11: Condition probabilities from the Hidden Markov Model output with five states ($AIC = 1964.374$).

	1 to 1	**1 to 2**	**1 to 3**	**1 to 4**	**1 to 5**
Child-Child	0.9379	0.0162	0.0148	0.0176	0.0128
Child-Robot	0.9185	0.0230	0.0163	0.0185	0.0222
	2 to 1	**2 to 2**	**2 to 3**	**2 to 4**	**2 to 5**
Child-Child	0.0210	0.9308	0.0159	0.0156	0.0158
Child-Robot	0.0174	0.9354	0.0187	0.0129	0.0145
	3 to 1	**3 to 2**	**3 to 3**	**3 to 4**	**3 to 5**
Child-Child	0.0218	0.0207	0.9140	0.0220	0.0204
Child-Robot	0.0256	0.0292	0.8991	0.0195	0.0263
	4 to 1	**4 to 2**	**4 to 3**	**4 to 4**	**4 to 5**
Child-Child	0.0149	0.0154	0.0155	0.9378	0.0158
Child-Robot	0.0263	0.0312	0.0307	0.8787	0.0313
	5 to 1	**5 to 2**	**5 to 3**	**5 to 4**	**5 to 5**
Child-Child	0.0165	0.0166	0.0158	0.0210	0.9298
Child-Robot	0.0225	0.0175	0.0215	0.0192	0.9185

Figure 12: Transition probabilities for an HMM with five states

References

Aarts, E. (2019a). Estimation of the multilevel hidden markov model.

Aarts, E. (2019b). Multilevel hmm tutorial.

Belpaeme, T., Baxter, P., De Greeff, J., Kennedy, J., Read, R., Looije, R., ... Zelati, M. C. (2013). Child-robot interaction: Perspectives and challenges. In *International conference on social robotics* (pp. 452–459).

Belpaeme, T., Baxter, P., Read, R., Wood, R., Cuayáhuitl, H., Kiefer, B., ... others (2013). Multimodal child-robot interaction: Building social bonds. *Journal of Human-Robot Interaction*, *1*(2), 33–53.

Beyreuther, M., Hammer, C., Wassermann, J., Ohrnberger, M., & Megies, T. (2012). Constructing a hidden markov model based earthquake detector: application to induced seismicity. *Geophysical Journal International*, *189*(1), 602–610.

Dautenhahn, K. (2007). Socially intelligent robots: dimensions of human–robot interaction. *Philosophical transactions of the royal society B: Biological sciences*, *362*(1480), 679–704.

Dean, B., Freeman, R., Kirk, H., Leonard, K., Phillips, R. A., Perrins, C. M., & Guilford, T. (2013). Behavioural mapping of a pelagic seabird: combining multiple sensors and a hidden markov model reveals the distribution of at-sea behaviour. *Journal of the Royal Society Interface*, *10*(78), 20120570.

de Haan-Rietdijk, S., Kuppens, P., Bergeman, C., Sheeber, L., Allen, N., & Hamaker, E. (2017). On the use of mixed markov models for intensive longitudinal data. *Multivariate behavioral research*, *52*(6), 747–767.

Eddy, S. R. (1998). Profile hidden markov models. *Bioinformatics (Oxford, England)*, *14*(9), 755–763.

Field, A. (2017). *Discovering statistics using ibm spss statistics, 5th edition.* London: SAGE Publications. Retrieved from `http://sro.sussex.ac.uk/id/eprint/77862/`

Fok, A. W., Wong, H.-S., & Chen, Y. (2005). Hidden markov model based characterization of content access patterns in an e-learning environment. In *2005 ieee international conference on multimedia and expo* (pp. 201–204).

Hamaker, E. L., & Wichers, M. (2017). No time like the present: Discovering the hidden dynamics in intensive longitudinal data. *Current Directions in Psychological Science*, *26*(1), 10–15.

Hox, J. J., Moerbeek, M., & Van de Schoot, R. (2018). *Multilevel analysis: Techniques and applications (3rd).* New York: Routledge.

Lemaignan, S., Edmunds, C. E., Senft, E., & Belpaeme, T. (2018). The pinsoro dataset: Supporting the data-driven study of child-child and child-robot social dynamics. *PloS one*, *13*(10).

Lynch, S. M. (2007). *Introduction to applied bayesian statistics and estimation for social scientists.* Springer Science & Business Media.

Mackenzie, A. (2010). *Wirelessness: Radical empiricism in network cultures.* MIT Press.

McHugh, M. L. (2013). The chi-square test of independence. *Biochemia medica: Biochemia medica*, *23*(2), 143–149.

Schafer, T. A. W. J. L. (2006). *Models for intensive longitudinal data.* Oxford University Press.

Shirley, K. E., Small, D. S., Lynch, K. G., Maisto, S. A., & Oslin, D. W. (2010). Hidden markov models for alcoholism treatment trial data. *The Annals of Applied Statistics*, 366–395.

Visser, I. (2011). Seven things to remember about hidden markov models: A tutorial on markovian models for time series. *Journal of Mathematical Psychology*, *55*(6), 403–415.

Visser, I., Raijmakers, M. E., & Molenaar, P. (2002). Fitting hidden markov models to psychological data. *Scientific Programming*, *10*(3), 185–199.

Wang, H.-C. (2008). Modeling idea generation sequences using hidden markov models. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 30).

Yu, W., & Sheblé, G. B. (2006). Modeling electricity markets with hidden markov model. *Electric Power Systems Research*, *76*(6-7), 445–451.

Zeger, S. L., & Liang, K.-Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, 121–130.