ABSTRACT

NEW METRICS AND MEASUREMENT FOR INFORMATION ACQUISITION IN DECISION MAKING

by Xiaolei Zhou

In recent decades, eye-tracking devices have become the widely used experimental tools to discover information acquisition strategies while human making decisions. However, the relevant metrics and measurements did not keep pace with devices’ improvements. Here, two metrics and their associated analytic methods that aim to capture the nature hidden in the eye movement sequences will be introduced. First, String Edit Distance (SED) will be used to assess the relative (dis)similarity between two different eye-tracking sequences, and reveal human behaviors patterns under different conditions. Second, the usage of Lag Sequential Analysis (LSA) will be proposed, which has the potential to describe the dynamic nature underlying such information acquisition eye movement data. To accomplish this goal, two associated metrics for LSA will be presented, the highest order of LSA, as well as its’ stationarity, which indicate whether the dynamic nature was consistent across the whole time course.
NEW METRICS AND MEASUREMENT FOR INFORMATION ACQUISITION IN DECISION MAKING

A Thesis

Submitted to the

Faculty of Miami University

in partial fulfillment of

the requirements for the degree of

Master of Arts

Department of Psychology

by

Xiaolei Zhou

Miami University

Oxford, Ohio

2014

Advisor: Dr. Joseph G. Johnson

Reader: Dr. Christopher Wolfe

Reader: Dr. Robin D. Thomas

Reader: Dr. Stephen Wright
# Table of Content

Introduction .......................................................................................................................... 1  
String Edit Distance (SED) Analysis .................................................................................. 7  
  String Edit Distance ........................................................................................................ 7  
  Needleman-Wunsch Algorithm & Scoring ......................................................................... 7  
  Substitution matrix .......................................................................................................... 10  
  The gap penalty ............................................................................................................. 10  
Experiment 1 ....................................................................................................................... 11  
  Method ............................................................................................................................ 11  
  Strategy-Tracing experiment ......................................................................................... 14  
Experiment 2 ....................................................................................................................... 22  
  Method ............................................................................................................................ 22  
  Time-pressure experiment ............................................................................................. 24  
  Lag Sequential Analysis (LSA) ....................................................................................... 32  
  Transition Matrix ........................................................................................................... 32  
  Markov Model ............................................................................................................... 36  
  Lag Sequential Analysis (LSA) ....................................................................................... 36  
  Internal dynamics of information acquisition sequences .................................................. 38  
Experiment 3 ....................................................................................................................... 41  
  Method ............................................................................................................................ 41  
  Results and Discussion ................................................................................................... 41  
  General Discussion ......................................................................................................... 44  
Reference ............................................................................................................................ 46  
Appendix A. String Edit Distance analysis procedure (NWD calculation) ...................... 48  
  Future direction: Significance level of NWD ................................................................. 50  
Appendix B. Lag Sequential Analysis procedure ................................................................. 51
List of Tables

Table 1. A simple example Substitution matrix ($S$) for four AOIs (A, B, C and D). A, B, C, D stands for 4 AOIs in a simple 2 x 2 information matrix design (2 options with 2 attributes). The values in the matrix stand for the costs when aligning every pair of AOIs (in different sequences). Scores were generated based on Euclidian distance between two AOIs, and will be used as the examples in the Appendix. Gap penalty is set as 0, the relative higher gap penalty is, and the more likely gaps will be inserted during sequence alignment process, relative lower gap penalty means substitutions would be favored other than gap insertions. ........................................................................................................... 9

Table 2. Data analysis plan for SED analysis on Strategy-Tracing experiments (6 attributes). The values in each cell are the mean NWD scores within same block across all trials and all participants; the numbers in the parenthesis is the 95% confidence interval for the mean NWD. Colored cells indicated that higher similarities for the same empirical sequence comparing to different theoretical sequences................................................................................................................................................. 20

Table 3. Data analysis plan for SED analysis on Strategy-Tracing experiments (3 attributes)..................................................................................................................................................... 21

Table 4. Descriptive statistics of Time pressure data. ITP is Increasing Time Pressure, DTP is Decreasing Time Pressure........................................................................................................................................... 25

Table 5. The stationary analysis table for the example sequence 'BABDCDBCDABABCAB'. ........................................................................................................................................................................ 40

Table 6. Sequence alignment of two example sequences. Used the substitution matrix above in table 2. All colored cells in the table show the back-tracking path to find out the optimized global alignment to transform ‘CGAGACGT’ into ‘AGACTAGTTAC’. Green colored items mean ‘match’, ‘-’ & grey colored items mean gap insertions, ‘|’ & yellow colored items mean substitutions by using substitution matrix......................................................... 49
List of Figures

Figure 2. Examples of four information acquisition strategies. We can use SI and SM to differentiate such two big categories of strategies (WADD & EQW vs. LEX & EBA), since SI and SM for WADD & EQW are positive, but negative for LEX & EBA. Even more, by using multiple-step transitions, the resulting proportions of different types of transitions also show the difference between two categories of strategies.............. 5

Figure 3. Sample stimuli used in strategy-tracing experiment. Three choices (Rows) with three attributes (columns) were created. 9 AOIs were labeled to ease data analysis, from (A) to (I). Numerical values were presented in each cell in the information matrix. .......................................................................................... 12

Figure 4. Sample stimuli used in strategy-tracing experiment. Three choices (Rows) with six attributes (columns). There are 18 AOIs in this sample stimuli, (A) to (R). Participants were asked to use either LEX or EQW to make their final decision. If they used LEX to acquire information in the table, they would throw away Apt A when they move acquisitions from ‘Value’ to ‘Location’, and then stay at Apt B and C only, until ‘Security’. Finally, one would pick Apt C. If one were asked to use EQW, participants will acquire all information row by row, and equally summarize all attributes of every apartment, and pick Apt A at last.................................13

Figure 5. Statistical mean (M) of number of acquisitions for each empirical and theoretical conditions. EQW and LEX are the strategies that participants were asked to apply when they were doing information searching experiments. ‘_3’ and ‘_6’ indicates whether the stimuli are of 3 attributes or 6 attributes. (Order A, B and Theoretical conditions could find in text.) ................................................................. 16

Figure 6. Comparisons of participants’ performances between Order A and Order B. Both conditions were compared to the associated theoretically generated sequences. For the participants in Order B, they applied strategy EQW (EQW_3 & EQW_6) better than participants in Order A. However, participants in Order A outperformed participants in Order B on applying strategy LEX (both LEX_3 and LEX_6). Also, whichever strategy applied last had higher similarities than other experimental blocks; this could be explained due to practice effect. ................. 18

Figure 7. Sample stimuli used in time pressure experiment. Three choices (Rows) with several attributes (columns) will be showed to all participants during the experiments. All the information for the current experiment will use binary codes (+ or -). Each cell’s information will be occluded unless an eye fixation is encountered on one cell. The choices’ and attributes’ labels will be available across the whole experimental processes........................................................................................................ 23

Figure 8. Number of acquisitions for two general conditions. ITP is Increasing Time Pressure, DTP is Decreasing Time Pressure................................................................. 27

Figure 9. Time per acquisition for two general conditions. ITP is Increasing Time Pressure, DTP is Decreasing Time Pressure................................................................. 28

Figure 10. SED comparison between ITP and DTP on 30s, 18s, and 6s. And SED comparison between ITP and DTP on 36s, 24, and 12s. The left figure used 36s as baseline, we can see higher time pressure impaired participants’ performances for
both conditions (ITP vs DTP). The right figure used 30s as baseline for SED comparisons. There was significant difference in information searching strategies between 18s and 6s. Moreover, the order of changing time pressure also caused significant difference on the performances, green bars under two conditions are significantly different.

Figure 11. First order (lag=1, digram: \((t-1,t)\) ) transition matrix of above example sequence. ‘A’, ‘B’, ‘C’, ‘D’, each stands for the eye fixations at one AOI. Subscript equals 0 means the first AOI in \((t-1,t)\) digram, subscript value equals 1 means the second AOI in \((t-1,t)\) digram. Grey cells are constructive zeros. Empty cells stand for zero frequency of corresponding transitions. For details on the creation of this example transition matrix see (Appendix. B).

Figure 12. Second order (lag = 2, trigram: \((t-2,t-1,t)\) ) transition matrix for the example information acquisition sequence. Grey cells are constructive zeros. Empty cells indicate zero frequency of relative trigram AOIs.

Figure 13. Increasing time pressure data analysis results by using LSA analysis. Y-axis in every sub-figure is all 9 trials for each participant, blue indicates zero order effect (random effect), green indicates first order effect. X-axis for each sub-figure is subject number, current analysis includes 21 participants.

Figure 14. Decreasing time pressure data analysis results by using LSA analysis. Y-axis in every sub-figure is all 9 trials for each participant, red indicates zero order effect (random effect), light blue indicates first order effect. X-axis for each sub-figure is subject number, current analysis includes 28 participants.

Figure 15. A fraction of one eye movement sequence, which indicate that subject looked over choice A across all attributes and associated first order transition matrix. Subscript 0 means the first cell in one digram. Subscript 1 means the second move in the same digram.
Acknowledgements

I cannot express enough thanks to my advisor for his advices and support: Dr. Joseph G. Johnson. Also I would like to express my gratitude to my committee for their valuable comments and encouragement: Dr. Christopher Wolfe, Dr. Stephen Wright, and Dr. Robin Thomas. I offer my sincere appreciation for the learning opportunities provided by my committee.
Introduction

Payne, Bettman and Johnson (1993) suggested that decision making strategies require different cognitive efforts and every decision made by human beings require searching and integration skills through a complex amount of information (different alternatives or attributes). In their book, the adaptive decision maker (Payne et al., 1993), they treated the decision maker as a limited-capacity information processor, and the decision making processes are highly contingent on the properties of decision tasks. However, they believe decision makers are highly adaptive under different decision making task circumstances. Since the nature of decision tasks is flexibility, such characteristic could have long-run benefits for decision makers, but also could cause short-run errors during decision making processes. So, if we switch the angle from evaluating final outcomes of decision making tasks to discovering the processes when people making decisions. In other words, rather than understanding whether such adaptive behavior would cause benefits or losses in decision makers’ lives, we can ask another more fundamental question, what are the rudimental cognitive processes used to make different decisions. Based on such perspective, Payne, et al (1993) proposed several decision making strategies that hypothetically captured different cognitive processes. The definition for such decision strategy is “a sequence of mental and effector (actions on environment) operations used to transform an initial state of knowledge into a final good state of knowledge where decision makers view the problem as solved” (p.9., Payne, Bettman & Johnson, 1993). As one can easily realize this definition pays attention to the cognitive aspect of different decision making processes.

To understand such decision making processes more deeply, researchers (Payne, 1976; Payne, Bettman & Johnson, 1988) used information cards (each information card will be treated as an AOI (Area Of Interest)), and more recently, applied the usage of computer mouse to track such processes that how people acquire information from AOIs (Areas that contain different information on computer screen) and make decisions. In order to discover participants’ information acquisition patterns, researchers would ask subjects either to flip over the information cards to see the relative information for one choice’s one attribute, or asked to move the mouse to uncover the occluded information cells on a computer screen, (e.g. Bettman, Johnson, & Payne, 1990; Payne, Bettman, & Johnson, 1993; Johnson & Koop, 2011), MouselabWeb (Willemsen & Johnson, 2011)

Moreover, for the recent four decades, eye tracking (Rayner, 1998; Duchowski, 2007; Russo, 2011) has become a widely used information acquisition tracing method. Since the normal sampling frequency of eye tracking devices is 50hz or 60hz (even higher, up to 2000hz), this produces a data matrix with a large amount of deliberate process sequences. By using such high-resolution sequencing data, we are facing a great opportunity to reveal the secret in ‘the black box’, that is, how human beings acquire information and mentally process it, through eye fixation sequences. (Payne, Bettman, & Johnson, 1993; Glöckner & Witterman, 2010; Horstmann, Ahlgrimm, & Glöckner, 2009; Brandstätter, Gigerenzer, & Hertwig, 2006; Krajbich, Lu, Camerer, & Rangel, 2012; Koop & Johnson, 2013).
However, the metrics deployed in analyzing such eye tracking data have not kept pace. Most of the information acquisition analyses in the field of decision making were executed on sequence of fixations to AOIs, often times descriptive statistics were reported (e.g. average fixation duration, number of fixations, proportion of information acquired, etc. See Duchowski, 2007 for a review). More recently, Ball (1997) and Russo (2011) used richer but still crude metrics (single-step fixation transitions and multiple-step fixation transitions) to analyze such data. In order to measure and identify the processes of different theoretical decision making strategies (e.g Equally weighted (EQW\textsuperscript{1}), lexicographic (LEX\textsuperscript{2}), etc) that Payne et al (1993) proposed, Ball (1997) proposed multivariate analysis method to capture the essence of different information acquisition processes of decision making strategies. Based on the strategy index (SI) (Eq.1) (type I to IV) that Payne (1976) introduced by using the relative frequency of two single-step transitions (alternative-wise vs. attribute-wise) in the information searching sequence, as well as another similar metric, strategy measure (SM) (Eq.2) proposed by Bockenholt and Hynan (1994).

\[
\text{Strategy Index (SI)} = \frac{N_{II} - N_{III}}{N_{II} + N_{III}} \quad (\text{Eq.1})
\]

\[
\text{Strategy Measure (SM)} = \sqrt{\frac{N_{\text{Total Transition}}}{(N_{alt}N_{att}/N_{\text{Total Transition}})(N_{II} - N_{III}) - (N_{att} - N_{alt})}} \quad (\text{Eq.2})
\]

They proposed a multivariate comparison, which other than considering only single-step transitions, multiple-step transitions were also taken into consideration as the measure of information acquisition strategies. For example, they (Bockenholt & Hynan, 1994) proposed 2-attribute comparison (type VI) and pair wise (2-choice) comparison (type V). Even further, a 3-attribute comparison (type VII) was also introduced (figure 1).

---

\textsuperscript{1} Equally weighted strategy: one will give equal considerations on every attribute of each option; the option with highest composited score will be picked up at last.

\textsuperscript{2} Lexicographic strategy: one will search for attribute cues, which would be used as flag to reduce the number of possible options that one could choose from.
Figure 1. Seven types of single-step transition that described by Ball, C. (1997). (Reproduced from Ball's paper.)
By using these proposed metrics, we had a chance to deepen our understanding of information acquisition strategies that reflected on eye-tracking data. Glöckner and Herbold (2011) used descriptive statistics of eye tracking data as well as transition analysis (relative proportion of different types of transitions (within gamble vs. between gamble)), to analyze information acquisition patterns associated with different decision making strategies in risky decisions. Also, Day et al (2009) used descriptive metrics and transition metrics to test the effects of music tempo and task difficulty on multi-attribute decision making tasks. Similarly, Horstmann, Ahlgrimm and Glöckner (2009) used SM-index as an indicator of direction of information search and information integration strategies. In sum, the analysis of descriptive statistics of eye-tracking data are necessary and easy to conduct, but we can see a clear trend of emerging eye fixation transitions analysis in finding information acquisition patterns for different decision making strategies.

Although useful, current transition metrics (e.g. SI and SM in Figure 2) are still crude, since they still only reflect partial information of the whole information acquisition sequences recorded by eye-tracking devices (Schulte-Mecklenbeck et al., 2011).
Figure 2. Examples of four information acquisition strategies. We can use SI and SM to differentiate such two big categories of strategies (WADD & EQW vs. LEX & EBA), since SI and SM for WADD & EQW are positive, but negative for LEX & EBA. Even more, by using multiple-step transitions, the resulting proportions of different types of transitions also show the difference between two categories of strategies.
It is logical to consider analyzing each information acquisition sequence as one large piece of information, which incorporated transitions, to further analyze information acquisition strategies. Luckily, there has been existing metrics and analyzing tools to work on such high-resolution data in other fields. Researchers in bioinformatics are using string edit distance on daily basis to compare the similarities between two or more DNA or protein sequences. Brandt and Stark (1997) have already used such string edit distance (Levenshtein, 1966) to analyze human visual searching patterns, and proposed Scanpath theory, which treated an eye movement sequence as one whole piece information, and comparing the similarities between different sequences. Moreover, such method has been adopted broadly and successfully in many other different fields, for example, human perception (Privitaria & Stark, 2000), fixation patterns (West et al, 2006), etc. Day (2010) and Cristino, Mathot, Theeuwes & Gilchrist (2010) have indeed deployed the use of SED in decision making. Day (2010) has made early explorations of applying such metric in multiple choice decision making tasks. He compared the similarities between theoretical (TISS – typical information search sequence) and empirical (EISS – empirical information search sequence) eye movement sequences. However, Cristitino et al (2010) designed a toolbox for MATLAB (ScanMatch toolbox), which give us an opportunity to easily use SED and further discover the most appropriate theoretical usage of such metric. Inspired by the above two explorations, I will use ScanMatch toolbox to analyze human information acquisition processes under different time pressure conditions, as well as on how precise human can do if they were asked to accomplish decision making tasks by using specific strategies (LEX & EQW). Beyond the previous works, the critical contributions of the current work are: developing further theoretical exploration on SED; testing the significance levels of (dis)similarities within and across levels of the experimental manipulations (i.e. different time pressure levels: 36s vs. 24s, and 24s vs. 12s, etc.), as well as comparative model evaluation (e.g. Theoretical-strategy derived sequences vs Empirical-strategy sequences, etc.).

Other than using SED to test the (dis)similarities between different information acquisition sequences, another important aspect of nature of such sequential data would be its internal dynamics that are hiding in the transition matrices. Gottman and Roy (1990) proposed lag sequential analysis (LSA) for behavioral researchers (e.g. group discussion analysis), which give us an opportunity to transport LSA into decision making research and unravel the trees in the forest of the different orders’ transition matrices. By using LSA we could detect how many previous steps of information acquired relate to the current one (the ORDER of LSA). Moreover, Bakeman & Quera (1995) published log-linear approaches to such sequential analysis (primarily hierarchical log-linear LSA)\(^4\). So, here I will propose to apply the log-linear LSA to information acquisition data under different time pressure (Koop & Johnson, 2012), and try to capture the internal dynamics of information acquisition strategies in the aspects as described above. Based on such sequential information of eye-tracking data, we will discuss the further use of such important information on exploring the underlying cognitive processes in decision making tasks—and beyond to other cognitive domains.

\(^3\) All the log-linear analyses in the current paper will use log represents nature log (logarithm to the base e).

\(^4\) Bakeman & Quera (1995) used likelihood ratio tests along with an iterative proportional fitting algorithm (Fienberg, 1970) to discover behavioral patterns; I will follow the suit in the current work.
String Edit Distance (SED) Analysis

In order to discover new metrics that will function beyond existing analysis on eye movement sequences, we will introduce a new metric in decision making, which adopted from Bioinformatics, the field has the analytic tool to analyze sequences similar to eye movement sequences (i.e. DNA sequences and protein structures), called String Edit Distance (SED). SED counts the editing costs of transforming one sequence into another, on the basic premise that more similar strings would have lower transformation costs (or less transformation operations). Commonly used transformation operations include insertion, deletion, match, substitution and gaps (described in detail below). By assigning values for different editing operations, as well as using dynamic programming techniques (e.g. Needleman-Wunsch Algorithm (NWA)) to optimally transform one sequence into another, and make the two mostly the same, then such global (dis)similarity between two sequences will be measured based on the ‘costs’ of all transformation operations. Finally, the similarity comparison between two eye movement sequences is accomplished.

String Edit Distance\(^5\)

Leveshtein distance (LD) (Levenshtein, 1966) is a widely applied metric, which measures the minimum costs of the transformation from one sequence into another, by using three basic operations (insertion, deletion, substitution), each having its own cost. As long as the alignment transformation was computed, LD shows the minimal alignment score, and indicates the similarity of two sequences.\(^6\) LD has accomplished our basic need to compare two sequences, however, on the other hand, three degrees of freedom in parameter settings (costs of insertions, deletions and substitutions) is a little bit too complicated. Luckily, an algorithm that has been used in bioinformatics for several decades, called the Needleman-Wunsch algorithm (NWA), requires only two degrees of freedom (substitutions and gaps: Day, 2010; Cristino et al., 2010). Even more, NWA can be combined with back-tracking optimization algorithms, which compute the similarities between two or more sequences automatically. Similarly, but not the same as LD, the higher the Needleman-Wunsch Distance (NWD) is, the more similar two sequences are.

Needleman-Wunsch Algorithm & Scoring

Needleman-Wunsch Algorithm (NWA) (Needleman & Wunsch, 1970) is a global alignment method that uses dynamic programming to optimize the string editing operations that transform one sequence into another. There are two major differences between NWA and LD, first, LD includes three editing operations, insertion, deletion and substitution, NWA includes only substitutions and gaps, which means, NWA has less parameter than LD, and this is a main reason of choosing NWA instead of LD for the current analysis. Second, LD is focusing on minimizing the number of editing operations, however, NWA is maximizing the alignment scores globally, by using back-tracking method. In other words, the lower the score of LD the more similar are two sequences--with no lower and upper boundaries, thereby hindering interpretation. However, the

\(^5\) From now on, when we talk about sequence, we also mean such sequence as a whole is one string, and one string equally means one sequence. The wording of string and sequence will use alternatively.

\(^6\) Technically, the higher the LD score is, the greater dissimilarity between two sequences.
NWA score is normalized by (Eq. 3), which provides the closed boundaries for NWA scores: [0,1]. Theoretically, the final score generated after normalization suggests a similarity metric defined by the range between 0 and 1, 0 means “two sequences are completely different”, 1 means “two sequences are exactly the same.”

\[
\text{Normalized NWD} = \frac{\text{Alignment Score}}{\text{Max(SubstitutionMatrix)} \times \text{Max(Length}_{1\text{st sequence}}, \text{ Length}_{2\text{nd sequence}})}
\]  

(Eq.3)

To practically compute NWD, first, we need to use ScanMatch toolbox in MATLAB (see Appendix A.) by specifying a substitution matrix that includes all the possible scores of each alignment operations (“costs”), one can treat these ‘cost’ values as ‘master keys’ when aligning two sequences of AOIs one by one. We also need to assign a gap penalty to allow for a cost of misaligning rather than substituting elements (simple example: Table 1). Second, we will use back-tracing method to select the optimal alignment operations. Finally, we normalize the editing costs (min[NWD]) relative to the length of longer sequence in the pair-wised comparison (Eq.3).
<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>3.5</td>
<td>2.5</td>
<td>2.5</td>
<td>2.0858</td>
</tr>
<tr>
<td>B</td>
<td>2.5</td>
<td>3.5</td>
<td>2.0858</td>
<td>2.5</td>
</tr>
<tr>
<td>C</td>
<td>2.5</td>
<td>2.0858</td>
<td>3.5</td>
<td>2.5</td>
</tr>
<tr>
<td>D</td>
<td>2.0858</td>
<td>2.5</td>
<td>2.5</td>
<td>3.5</td>
</tr>
</tbody>
</table>

Gap penalty = 0

Table 1. A simple example Substitution matrix ($S$) for four AOIs (A, B, C, and D). A, B, C, D stands for 4 AOIs in a simple 2 x 2 information matrix design (2 options with 2 attributes). The values in the matrix stand for the costs when aligning every pair of AOIs (in different sequences). Scores were generated based on Euclidian distance between two AOIs, and will be used as the examples in the Appendix. Gap penalty is set as 0, the relative higher gap penalty is, and the more likely gaps will be inserted during sequence alignment process, relative lower gap penalty means substitutions would be favored other than gap insertions.
**Substitution matrix**

In bioinformatics and evolutionary biology, a substitution matrix \((S)\) describes the rate at which one character in a sequence transform into other character states over time, where higher score indicates higher similarity, and generically speaking, less effort would be needed to put through while applying such substitution during sequence alignment. By taking account of this feature, we can take the advantage of this more meaningful and simpler setting. Other than LD, which has three different operations instead of only one for NWD, one could have the power to consider the physical relations of AOIs’ as formalized in \(S\). A the same line, the substitution values themselves might also shed light on the mental effort of different movements that reflected in the substitution matrix.

There are different ways to create \(S\); basic intuition as well as the \(S\) used for this paper is to simply assuming a positive relationship between \(S\) values and the physical distance between different AOIs to establish the substitution matrix. Specifically, we compute the Euclidean distance between each pair of AOIs and inversely insert into substitution matrix, which the highest score means the identical match of two characters, the lowest score means the furthest apart. Not only could one easily employ other distance metrics where appropriate (e.g., City Block distance if diagonal transitions are infrequent or impossible), but also incorporate other theoretical relationships (e.g. semantic relationships, etc) among two consecutive AOIs in forming \(S\).

**The gap penalty**

The last important parameter for NWA is the gap penalty. Its value is directly reflected in sequence alignment. A small or negative gap penalty leads for favoring more substitutions in the sequence alignment processes, other than gap insertions. Reversely, a relative large gap value (comparing to other substitution costs in the substitution matrix) of inserting a gap would make NWA more likely to apply gap insertions rather than substitution when choosing the operations (substitution and gap insertion) to align the paired AOIs in different sequences. Moreover, the inserted gaps could be potentially interpreted as the clustering “flags” to justify different cognitively meaningful subsequences. In sum, gap penalties could be set up beforehand if some hypothesized \(S\), or could be optimized by practically testing different gap penalty values based on \(S\).

The following two experiments will demonstrate how to apply SED on the comparisons between empirical sequences themselves, as well as empirical sequences and theoretically generated eye movement sequences.
Experiment 1

In this experiment, we will assess the contributions of SED analysis to the analysis of information acquisition strategies by comparing empirical information acquisition mouse-tracking sequences to theoretically hypothesized information acquisition strategies’ sequences under the settings of preferential decision making tasks (3 options X (3 or 6) attributes). (Johnson & Koop, 2011). The NWD will be generated by using the ScanMatch toolbox (Cristino, et al., 2010) in MATLAB.

Method

The following is the setup of a classic preferential decision making task, for the current strategy-tracing experiment, 3 options were picked up, and 3 and 6 attributes stimuli were used, respectively. (Figure 3 & 4)
Figure 3. Sample stimuli used in strategy-tracing experiment. Three choices (Rows) with three attributes (columns) were created. 9 AOIs were labeled to ease data analysis, from (A) to (I). Numerical values were presented in each cell in the information matrix.

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
<th>Location</th>
<th>Amenities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apt A</td>
<td>(A) 6</td>
<td>(B) 9</td>
<td>(C) 6</td>
</tr>
<tr>
<td>Apt B</td>
<td>(D) 8</td>
<td>(E) 5</td>
<td>(F) 6</td>
</tr>
<tr>
<td>Apt C</td>
<td>(G) 8</td>
<td>(H) 5</td>
<td>(I) 7</td>
</tr>
</tbody>
</table>
Figure 4. Sample stimuli used in strategy-tracing experiment. Three choices (Rows) with six attributes (columns). There are 18 AOIs in this sample stimuli, (A) to (R). Participants were asked to use either LEX or EQW to make their final decision. If they used LEX to acquire information in the table, they would throw away Apt A when they move acquisitions from ‘Value’ to ‘Location’, and then stay at Apt B and C only, until ‘Security’. Finally, one would pick Apt C. If one were asked to use EQW, participants will acquire all information row by row, and equally summarize all attributes of every apartment, and pick Apt A at last.
**Strategy-Tracing experiment**

*Participants.* Seventy-seven introductory psychology students from a mid-sized Midwestern U.S. university participated in the study. Participants selected the study from an online sign-up site, where this particular study was listed amongst many others. In return for their participation, all participants earned course credit as well as a performance contingent payment.

*Materials.* In the main task, participants were required to make choices between sets of three hypothetical apartments. The stimuli we used in the current study were designed to test how well that participants could apply two specific information acquisition strategies, LEX and EQW strategies. We also designed two big categories of stimuli to further test such information acquisition process: 3 attributes and 6 attributes for different options, respectively.

*Number of attributes.* The apartments used in this task were described by either three or six attributes. Trials with only three attributes (Figure 3) described the apartments with regards to “value,” “location,” and “amenities,” whereas trials with six attributes (Figure 4) also included “security,” “tenants,” and “landlord (Koop & Johnson, 2012).” The presentation order of these attributes across columns was consistent across all trials. Values on these attributes were provided for each apartment using a scale from 1 to 10, where higher numbers signified better values on each of the dimensions.

*Procedure.* When the participants arrived at the testing room, they were provided informed consent forms from experimenters. After provided with brief introduction and been noticed that they would be paid based on performance, experimenters oriented them logging on the computers and watched detailed instructions about the upcoming task. All participants were told to apply specific strategies when making decisions, although there were not yet given the details of any strategies.

For all participants, they were randomly assigned to two groups, Order A (42 participants in total) and Order B (35 participants in total). For both groups, participants were asked to complete 5 blocks of tasks, block 1, 3 were practice blocks, block 2 and 4 were real task blocks, block 5 was a surprise memory recall test block. For participants in the group of Order A, they were asked to deploy EQW strategy to complete all trials in the first experimental block (block 2), then use LEX strategy to complete all trials in the second experimental block (block 4). For all the experimental blocks, they faced 9 trials with 3 options X 6 attributes first, then another 9 trials with 3 options X 3 attributes. For the group of participants in Order B, they faced same practice blocks and experimental blocks, however, they were asked to apply LEX strategy to complete first experimental block, and EQW to accomplish the second one. In the task, each information cell was occluded unless mouse moved to such. We used process-tracing method to record their mouse movements for the following data analysis.

*Results and Discussion.* In the current experiment, each participant was asked to apply two strategies and a total 36 empirical tests (18 trials were accomplished by using LEX, and another 18 trials by using EQW). For every 18 trials in the same category, there were 9 trials designed with 3 options X 6 attributes, and 9 trials with 3 options X 3 attributes. Each participant was randomly assigned into two big experimental groups (Order A and Order B). 42 participants completed the task under Order A and 35 completed under Order B. In order to compare participants’ performances, graduate
students used same strategies to complete same experiments with the same order to generate associated mouse movement sequences as KEYS, then these key sequences were treated as theoretical sequences, and further compared with empirical sequences to discover the wellness of participants behaviors of applying specific information searching strategies.

As one can see from the following descriptive statistic (figure 5) of research conditions: Order A and Order B, first of all, generally speaking, it is easy to generate, also it is the most widely used and most basic metric to analyze mouse tracking data. In figure 5, it is logical that theoretical EQW sequences have higher number of total acquisitions than theoretical LEX sequences, as well as conditions with more attributes, which are both consistent with cost-benefit theory (Beach & Mitchell, 1978; Payne, Bettman, & Johnson, 1993; Chu & Spires, 2003). EQW need more effort to apply, however, LEX would need less. And this is really the case that showed in the summary statistic (figure 5) for Order A (no matter how many attributes in there). However, for Order B such theory did not hold steady practically. In the condition of Order B, where participants were asked to apply LEX first then EQW, for 6-attribute stimuli, started showing insignificant differences between EQW_6 and LEX_6. Moreover, the hypothesis even got reversed, for 3-attribute stimuli cases; LEX_3 has higher number of acquisitions than EQW_3. So, in other words, the selecting order of different decision making strategies would compromise the assumption that made by cost-benefit theory, which EQW would need more effort other than LEX strategy. Even more, from the figure when the participants were asked to use EQW, it shows that total numbers of acquisitions are higher with more attributes. But, if we are talking about same measurement for LEX, it seems the number of attributes did not cause any effects, where one possible explanation of this phenomenon could be, for compensatory strategies (i.e. EQW), there might be an interaction between accuracy and number of attributes, however, for non-compensatory strategies (i.e. LEX), such interaction might not exist, or could has been eliminated by the nature of these decision strategies. In sum, from this summary statistic, it seems that we found a pinch of evidence where there could be an interaction between the effort and accuracy of decision strategies, but this only piece of information is not adequate for us to draw any conclusion.
Figure 5. Statistical mean (M) of number of acquisitions for each empirical and theoretical conditions. EQW and LEX are the strategies that participants were asked to apply when they were doing information searching experiments. ‘_3’ and ‘_6’ indicates whether the stimuli are of 3 attributes or 6 attributes. (Order A, B and Theoretical conditions could find in text.)
However, by using SED analysis, we can perform richer analysis for the current experiment. For the first step of SED analysis, we will transform stepwise information acquisition sequences of AOIs (i.e. A B C D ...) into one whole sequence (AaBbCcDd...), which means we could take all transitions between every single steps’ information acquisition into consideration, not isolating each piece of information. By this step, we preserved transitional information between different steps’ acquisitions. Next, we need to set up the parameters for ScanMatch toolbox, substitution matrix and gap penalty. For this experiment, Euclidean distance is used to create the substitution matrix, and gap penalty was set up as 0. The Euclidean distance was calculated based on the distance between different physical centers of AOIs, but not the distance in pixels. Gap penalty was preset equals to zero, which means we did not favor to insert more gaps, we were more like to see how one sequence transform into another without inserting too many gaps while aligning two sequences. Finally, for the current research paradigm is to pass empirical sequences and theoretical sequences into ScanMatch toolbox to generate NWDs, then edit distances (similarity scores) between empirical sequences and theoretical one were computed. For example, in order to analyze the similarities between theoretical sequences and empirical sequences collected from the first 9 trials of all participants in block 2 of Order A (EQW_6 of Order A in figure 6), where the stimuli were 3 options X 6 attributes and participants were asked to use EQW strategies to complete all 9 trials. In order to ultimately compute the average similarity of sequences between empirical and theoretical in block 2 (Order A EQW_6), first, NWDs of all participants’ experimental sequences of trial 1 versus theoretical sequence of trial 1 were computed, and then the rest 8 trials’ NWDs were generated. Next, all 9 trials’ similarity comparisons were averaged out, and shown as one column in figure 6, named ‘EQW_6’. Other columns in figure 6 are created under the same procedure. As we discussed above, similarity scores generated by ScanMatch toolbox is between 0 and 1, 1 means two sequences (conditions) exactly same, 0 means two sequences (conditions) are completely different. In figure 6, we can see the similarities between empirical and theoretical conditions very clear. First, if we ignore the difference caused by Order A & B, in general, the information acquisition sequences which generated by asking participants to apply EQW have higher similarity of theoretical EQW sequences, other than LEX conditions. Which give us a very interesting result, for cost-benefit theory, EQW would need more effort but with higher accuracy, so theoretically speaking EQW should acquire more mental efforts to deploy, and LEX would be little bit relatively easier to apply, since LEX would require less mental effort theoretically. However, based on the SED analysis we could say EQW has relative higher accuracy than LEX, but LEX seems more difficult to practically apply when comparing to EQW. Moreover, there seems to be an ordering effect existing, since no matter the attributes are, the later strategies got deployed more similar to the theoretical ones. LEX got higher similarities than EQW did for Order A and EQW got higher similarities than LEX condition for Order B. To sum up, by doing SED analysis, one could incorporate more information other than just descriptive metric, and we can easily compare the participants’ behavioral wellness between different conditions.
Figure 6. Comparisons of participants’ performances between Order A and Order B. Both conditions were compared to the associated theoretically generated sequences. For the participants in Order B, they applied strategy EQW (EQW_3 & EQW_6) better than participants in Order A. However, participants in Order A outperformed participants in Order B on applying strategy LEX (both LEX_3 and LEX_6). Also, whichever strategy applied last had higher similarities than other experimental blocks; this could be explained due to practice effect.
At last, we could compare each empirical strategy with all theoretical strategies to see how well such participant applied on specific strategy other than all other theoretical strategies. So for example, when comparing Order A’s LEX_6 with both theoretical LEX_6 and EQW_6 (table 2 & 3), as shown in the table, the mean similarity between LEX_6 empirical and theoretical is 0.731, however, LEX_6 empirical vs EQW_6 theoretical’s similarity score is 0.314, which indicated that all participants in Order A applied LEX_6 pretty well.
<table>
<thead>
<tr>
<th>6 Attributes</th>
<th>Theoretical Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EQW</td>
</tr>
<tr>
<td>Empirical</td>
<td></td>
</tr>
<tr>
<td>Sequence</td>
<td>OrderA</td>
</tr>
<tr>
<td></td>
<td>EQW</td>
</tr>
<tr>
<td>OrderB</td>
<td>LEX</td>
</tr>
</tbody>
</table>

Table 2. Data analysis plan for SED analysis on Strategy-Tracing experiments (6 attributes). The values in each cell are the mean NWD scores within same block across all trials and all participants; the numbers in the parenthesis is the 95% confidence interval for the mean NWD. Colored cells indicated that higher similarities for the same empirical sequence comparing to different theoretical sequences.
<table>
<thead>
<tr>
<th>3 Attributes</th>
<th>Theoretical Sequence</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EQW</td>
<td>LEX</td>
<td></td>
</tr>
<tr>
<td>Empirical</td>
<td>OrderA</td>
<td>EQW</td>
<td>0.664, (0.593,0.694)</td>
</tr>
<tr>
<td>Sequence</td>
<td>LEX</td>
<td>0.507, (0.492,0.522)</td>
<td>0.767, (0.714,0.820)</td>
</tr>
<tr>
<td></td>
<td>OrderB</td>
<td>EQW</td>
<td>0.728, (0.678,0.779)</td>
</tr>
<tr>
<td></td>
<td>LEX</td>
<td>0.499, (0.471, 0.527)</td>
<td>0.654, (0.579,0.728)</td>
</tr>
</tbody>
</table>

*Table 3. Data analysis plan for SED analysis on Strategy-Tracing experiments (3 attributes).*
Experiment 2

In this experiment, we compare similarities between different task conditions (i.e., different time pressure levels), rather than look up the differences between theoretical models and empirical data.

Method

For this experiment, we will use classic multi-attribute multi-choices decision making tasks with binary codes. (Figure 7)
Figure 7. Sample stimuli used in time pressure experiment. Three choices (Rows) with several attributes (columns) will be showed to all participants during the experiments. All the information for the current experiment will use binary codes (+ or -). Each cell’s information will be occluded unless an eye fixation is encountered on one cell. The choices’ and attributes’ labels will be available across the whole experimental processes.
**Time-pressure experiment**

**Participants.** 49 introductory psychology students from a mid-sized Midwestern U.S. university participated in the study. Participants selected the study from an online sign-up site, where this particular study was listed amongst many others. In return for their participation, all participants earned course credit as well as a performance contingent payment.

**Procedure.** In the main task, after they signed informed consent forms, participants were required to make decisions between sets of three hypothetical movies. Each movie was presented in four characteristic dimensions, stars, budget, rating, and originality. And participants were asked to make decisions under different time pressure settings. Every trial was completed on computers, and the participants’ eye movements were recorded for the proposed analysis methods. Before they left the research room, credits were granted and all participants were debriefed by research assistant.

**Time Pressure & Design.** This time pressure (TP) study followed by 2 X 6 factorial design, (increasing/decreasing TP X 6 time slots). There were two general conditions, increasing time pressure (ITP), and decreasing time pressure (DTP). Under each condition, there were six different time pressure settings, which included six blocks, 6s, 12s, 18s, 24s, 30s, and 36s. For each block, there were 2 practice trials and 9 experimental trials presented to the subjects. For increasing time pressure condition, subjects faced 36s block first, with 6s as the increment, until accomplish the last 9 trials in 6s, and vice versa for decreasing time pressure. We created two big sets of 4 attributes * 3 alternatives stimuli, each set of stimuli included 9 different ones, these 9 stimuli were design to fulfill the permutation of 3 alternatives’ information across all 4 attributes. We used the first stimuli set on DTP block 1, 3, 5 and ITP block 2, 4, 6, and we used the second stimuli set on DTP block 2, 4, 6 and ITP block 1, 3, 5. Each stimulus under in one block was randomly presented to the participants.

**Results and Discussion.** For the current research paradigm as stated above, each participant faced 9 trials under every block, so every participant completed 54 trials in total (6 time slots X 9 trials). Since we want to show the advantage of our current analytic tool and new metric for process tracing data, so first I ran the descriptive analysis on number of acquisitions and time per acquisition. (Table 4.)
<table>
<thead>
<tr>
<th></th>
<th>Number of Acquisitions</th>
<th>Time per Acquisition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITP</td>
<td>M=82.58, SD=5.73</td>
<td>M=38.1, SD=13.56</td>
</tr>
<tr>
<td>DTP</td>
<td>M=84.92, SD=4.84</td>
<td>M=38.1, SD=11.46</td>
</tr>
</tbody>
</table>

Table 4. Descriptive statistics of Time pressure data. ITP is Increasing Time Pressure, DTP is Decreasing Time Pressure.
If we take a deeper look at participants’ performances (Figure 8 & 9). For the number of acquisitions, with more time (i.e. from 6s to 18s), participants would use more fixations to search information. (Figure 8). However, for time per acquisition we did not see large difference between different conditions. (Figure 9).
Figure 8. Number of acquisitions for two general conditions. ITP is Increasing Time Pressure, DTP is Decreasing Time Pressure.
Figure 9. Time per acquisition for two general conditions. ITP is Increasing Time Pressure, DTP is Decreasing Time Pressure.
Again, with only descriptive statistics of process tracking data, it’s hard to illustrate whether participants used different strategies under different time pressure conditions. So I used SED tool to generate NWD and compare participants’ performances under same conditions. (Figure 10.). In order to compare information acquisition similarities between different blocks, we need to choose “baseline”, and then compare blocks’ similarities relative to ‘baseline’. Since we designed two stimuli set, then we will compare the results separately. We set up ‘36s’ as ‘baselines’, then compared ‘12s’ and ‘24s’ in ITP and DTP. We can see ‘24s’ have higher similarity relative to ‘36s’, and ITP have higher similarity scores than same comparison in DTP. Similarly, we used ‘30s’ as another ‘baseline’, and compared ‘18s’ and ‘6s’. In this part of SED analysis we found some interesting results, but still, ‘18s’ have higher similarity other than ‘6s’ blocks for both ITP and DTP. However, for ‘6s’ block, DTP have higher similarity than ITP’s ‘6s’ block. In other words, when we are using ‘36s’ as baseline, we calculated the similarities between moderate time pressure (‘24s’) and no/low time pressure (‘36s’) and similarities between high time pressure (‘12s’) and no/low time pressure (‘36s’). From (A) in figure 10, we can see that ITP condition has generally higher similarities than DTP condition. And Moderate TP and High TP doesn’t show big differences for both ITP and DTP conditions. For (B) in figure 10, moderate TP and high TP also doesn’t show large discrepancies, however, for ITP there is a huge drop of similarity between moderate TP and high TP. One explanation for this could be humans seems more comfortable with adapting decision strategies while dealing with time pressure decreasingly. However, when dealing with increasing time pressure situations, it is possible that one might suddenly choke under high time pressure (Beilock & Carr, 2001; DeCaro, Thomas, Albert, & Beilock, 2011), and lose track of the strategies they used under no/low time pressure.
Figure 10. SED comparison between ITP and DTP on 30s, 18s, and 6s. And SED comparison between ITP and DTP on 36s, 24, and 12s. The left figure used 36s as baseline, we can see higher time pressure impaired participants’ performances for both conditions (ITP vs DTP). The right figure used 30s as baseline for SED comparisons. There was significant difference in information searching strategies between 18s and 6s. Moreover, the order of changing time pressure also caused significant difference on the performances, green bars under two conditions are significantly different.
In conclusion, if we put the analysis of descriptive statistics and SED analysis together, we have more information about participants’ information acquisition strategies under different time slots across increasing and decreasing time pressure conditions. With only descriptive statistics, we can only learn little knowledge about information acquisition, and it’s hard to incorporate different descriptive statistics into one metric to draw conclusions on decision making strategies. However, from figure 10, we used NWD as a new metric to retrieve the strategy similarities, which considered all information as a whole, it could generate one value to reflect different conditions’ information acquisition similarities.
Lag Sequential Analysis (LSA)

Payne et al (1993) summarized several decision making strategies (e.g. elimination-by-aspect (EBA), weighted additive (WADD), etc.), which we believe most of these strategies that adapted by decision makers are sequentially associated. Nowadays, although we can record high-resolution eye-tracking data, we don’t have an appropriate metric that could capture precisely about such sequential relationship. Intuitively, Markov analysis (Markov, 1971) should be considered as the ideal metric for such "gap," since the final order of Markov analysis can be used to detect the internal dynamics inside eye-tracking data, which we believe human information acquisition strategies hiding behind. However, because of the nature of decision making experimental paradigms, the transition matrices are often sparse (more than half of the entries in the matrix are zeros), then normal Markov analysis would fail, since the requirement for Markov analysis is that zero entries in the transition matrices should be no more than five entries. But luckily, several researchers studying social interaction (Bakeman & Gottman, 1986; Gottman & Roy, 1990) developed LSA, which gets around the deficit of Markov analysis in behavioral experiments, but kept the potential to capture the internal dynamics (order, stationarity, homogeneity) of behavioral sequences. Recently, Bakeman and Quera (1995, 2011) applied such method in social psychology, and behavioral sciences successfully. So, here, I will introduce this analytic tool, Lag Sequential Analysis (LSA), into decision making research field to discover the dynamics beneath eye fixation sequences occurring during the choice task. The highest order of lag could be another potential metric that measuring different information acquisition strategies in decision making research field.

Transition Matrix

Essentially, LSA is a hierarchical statistic model testing method based on the different orders’ of transition matrices. However, in order to generate transition matrix we need to introduce The Moving Time Window first, then discuss the creation of transition matrix for different orders (lags).

The Moving Time Window (TMTW). TMTW is a technique to create different orders’ transition matrices. The basic idea of TMTW is dividing the whole sequence into overlapping subsequences based on the smallest ‘time window’ defined beforehand (e.g. digrams (t-1,t) would produce a first-order matrix; trigrams (t-2,t-1,t) would produce a second-order matrix; etc). ‘Time window’ could be based on the frequency of eye-tracking devices, or it could be two different adjacent AOIs (1st order, digrams, (t-1,t)), or three different AOIs (2nd order, trigrams, (t-2,t-1,t)), or even more. We then tally the frequencies of all the transitions within each ‘time window’ to create the associated order’s transition matrix.

TMTW & Transition Matrix. For this part, we will discuss how to use TMTW and create different associated orders’ transition matrices. (e.g. figure.11). First, we will use TMTW, and create first order transition matrix (lag=1, digram: (t-1,t)) a matrix reflect the relationship between lag0 and lag1). Zero order transition matrix (lag=0) will only compute the frequencies of acquisitions for each AOI, which assumes structurally that there is no ordering effect at all. For example, assume a simple acquisition sequence containing only 4 AOIs, {A, B, C, D}:
First, we will create a 4-by-4 matrix with all zeros. Second, apply TMTW of digram: \((t-1,t)\) to the sequence, where the window is illustrated by use of parentheses for the first three digrams, and final digram, below:

\[
\begin{align*}
(BA)BDCDBCDABABCAB \\
B(AB)DCDBCDABABCAB \\
BA(BD)CDBCDABABCAB \\
\ldots
\end{align*}
\]

(For more detail see Appendix B). Next, for each diagram shown, add one to the relative cell in the transition matrix, where the cell row corresponds to the first letter of one AOI pair in the digram (e.g. B in the (BA)…; A in the B(AB)…), and the cell column is the second letter in such digram (e.g. A in the (BA)…; B in the B(AB)…). When repeated for each digram, a first order transition matrix is created (Figure 11). It is also possible to create higher order transition matrices (second order transition matrix or even higher, Figure 12, etc.)
Figure 11. First order (lag=1, digram: (t-1,t) ) transition matrix of above example sequence. ‘A’, ‘B’, ‘C’, ‘D’, each stands for the eye fixations at one AOI. Subscript equals 0 means the first AOI in (t-1,t) digram, subscript value equals 1 means the second AOI in (t-1,t) digram. Grey cells are constructive zeros. Empty cells stand for zero frequency of corresponding transitions. For details on the creation of this example transition matrix see (Appendix. B).
Figure 12. Second order (lag = 2, trigram: (t-2,t-1,t) ) transition matrix for the example information acquisition sequence. Grey cells are constructive zeros. Empty cells indicate zero frequency of relative trigram AOIs.
**Markov Model**

As long as we have different orders’ transition matrices in hand, it is logical to think about using Markov Model to determine the highest order significant transition matrix, because by using Markov Model analysis we can test the significance levels of internal dynamics (highest order, stationarity, homogeneity) that is reflected in the transition matrices. However, it’s not easily applicable to such transition matrices due to the current classical decision making experiments. To understand why, note that if we take a deeper look at the transition matrix (Figures 11, 12), we will see that many cells in the transition matrices are zeros. With this unique characteristic of current transition matrices, the minimum requirements for Markov Model analysis is not typically met, especially for a combination of many AOIs (i.e., attributes and/or options) and high-order matrices. However, Bakeman and Quera (1995) created a log-linear approach to analyze the transition matrix, which is similar to Markov Model, but gets around this deficient characteristic. In essence, they proposed to use hierarchical log-linear model testing to determine the highest significant order transition matrix, and then determine the highest ordering effect shown in the sequences.

**Lag Sequential Analysis (LSA)**

In general, Lag Sequential Analysis (LSA) is a hierarchical log-linear approach to determine highest order (largest lag in LSA) transition matrices of different trials, blocks, conditions and participants. To accomplish this goal, LSA will need to generate expected transition matrices for different orders under different statistical models (e.g independent models, conditional dependent models, etc.; for more detail see section ‘Model Testing’). Next, by using Likelihood Ratio Chi-Square tests (LRX²), LSA will test the significance level between two hierarchically-related models.

*Likelihood Ratio Chi-square (LRX²).* The basic statistic for LSA is the difference between observed transition frequency matrix and expected transition frequency matrix fitted by different models (see Model Testing), and a model’s goodness-of-fit will be tested by using Likelihood Ratio Chi-Square (LRX²; Eq. 5).

\[
G^2 = 2 \sum \frac{(OBS)_i}{(EXP)_i} \left( \frac{(OBS)_i}{(EXP)_i} - 1 \right) 
\]  

(Eq.5)

By using Equation 5, we can test different nested models hierarchically.

*Expected Frequency Table and IPF.* Like we discussed above, since the observed transition matrices are mostly sparse matrix (zero entries more than half of total cells in transition matrix), so we will use Iterative Proportional Fitting (IPF; Fienberg, 1970) algorithm to compute expected frequency table under different statistical models. Moreover, IPF can not only deal with first order expected transition matrices, but also can be used to generate higher order expected transition matrices for different models (see Fienberg, 1970 for additional details).

*Model Testing & Order of LSA.*

*LSA at Lag 1 (digram:(t-1,t)).* LSA analysis at lag = 1 is designed to disentangle the association between lag = 0 (AOI at (t-1) in digram: (t-1,1)) and lag = 1 (AOI at (t) in digram: (t-1,t)), which could give us the evidence to show the internal dynamics of information acquisition sequences by testing the relationship between current step’s AOI.

---

*Markov Model analysis requires that at least half of the cells in the transition matrix should not be zeros.*
and one previous step’s AOI: does the current focus of attention at any given moment during decision making depend on the previous focus of attention? Specifically, LSA at lag = 1 is hierarchically testing different statistical models (log-linear models) by fitting first-order observed transition matrix (using IPF). For LSA at lag = 1, two log-linear models will be tested. To be easy to discuss different model testing, we will adopt the common convention of bracket notation (Fienberg, 1970), but use it in a more intuitive way. For LSA at lag = 1, the first model is the saturated model denoted [t-1, t]; the second model is the independent model written [t-1][t]. By fitting different models, LSA will have two expected transition matrices, under each log-linear model ([t-1, t]-model and [t-1][t]-model; the difference between the fit of these two models is compared using Eq.6)

\[
\Delta G^2_{(t-1,t)} = G^2_{[t-1][t]} - G^2_{[t-1,t]} \tag{Eq.6}
\]

Since [t-1,t]-model is the saturated model, \(G^2_{[t-1,t]}\) will always be zero. We only need to calculate \(G^2_{[t-1][t]}\) by fitting [t-1][t]-model, then determine the significance level of such difference between two models (see Appendix B).

**LSA at Lag L (L ≥ 2).** If we find any significant difference between two first-order models (the [t-1,t]-model and [t-1][t]-model from above), then we can move on to higher-order LSA, starting from lag = 2, and repeat this until no significant difference is found between all the relevant log-linear models at Lag = L (see Appendix B). For example, we need to hierarchically test three models at Lag = 2 as follows. First, follow the same procedure of **LSA at Lag =1**, we will use IPF to fit reduced model 1, so that we can test a second order effect indicating a homogeneous association between lags t-2, t-1, and t. Since [t-2, t-1, t] is the Saturated Model, so \(G^2_{[t-2,t-1,t]}\) will always be 0, only \(G^2_{[t-2,t-1][t-1,t][t-2,t]}\) needs to be computed, then we can test its significance level. If we find a significant Lag = 2 relationship, we can move on to Lag=3 level LSA; otherwise, we will stay at Lag=2, and test the difference between reduced model 1 and reduced model 2, which can be used to test the conditional second-order association. From last step, we get \(G^2_{[t-2,t-1][t-1,t][t-2,t]}\) for the current step we need use IPF to fit [t-2,t-1][t-1,t]-model, and calculate \(\Delta G^2_{(t-2,t-1,t)}\) then we can test the significance level of the difference between two reduced models, as before (Equation 7):

\[
\Delta G^2_{(t-2,t-1,t)} = G^2_{[t-2,t-1][t-1,t][t-2,t]} - G^2_{[t-2,t-1,t]} \tag{Eq.7}
\]

If a significant difference is found, then we can say there is a second-order relationship found, and will move on to higher-order LSA (Lag = 3); if there is no significant effect at all, then we will conclude that the highest ordering effect in such information acquisition sequence is first-order (lag=1). To extrapolate under the same logic, if significant effect can be found for lag=L (L = 3, 4, ... etc), then we will move to (L+1) order LSA, if nothing significant for lag=L, then we will claim the highest ordering effect for the current information acquisition sequence is (L-1)-order.
**Internal dynamics of information acquisition sequences**

Recalls the reason that we want to apply LSA to the information acquisition sequences is to detect the internal dynamics of these sequences. Now that we know the basics of how to use LSA to test one sequence step by step, then we can get back to our initial goal, and discuss about how to obtain these core dimensional measures of information acquisition.

**Order.** From previous section, we already know how to determine the highest significant ordering effect by using LSA at different lag. As long as we find out the highest significant lag, for example, lag = 1, then we can say, for this trial, the information searching strategy that participant used in this trial is depending on one previous step’s information. If lag =2, then it suggests that participants used three AOIs’ information at every single mental process toward to the decision they made.

**Stationarity.** Another index that reflects the internal dynamic structure of information acquisition sequence would be the stability of different types of transitions over time in every trial. In practice, rather than including every single possible transition (i.e., cells in the transition matrix), we will categorize these transitions into the three single-step transition types proposed by Ball (1997; Type II, III, and IV). We want to have the power to test whether the probabilities of these three transitions (figure 1) in the first half of one trial stayed the same as in the second one. We will use likelihood ratio test (LRX²) to conduct the omnibus test for stationarity (Eq. 8).

\[
LRX^2 = 2 \sum_{t=1}^{T} \sum_{i_0, i_1, i_2, \ldots, i_r} n_{i_0, i_1, i_2, \ldots, i_r}(t) \log \frac{p_{i_0, i_1, i_2, \ldots, i_r}(t)}{\hat{p}_{i_0, i_1, i_2, \ldots, i_r}}
\]  
(Eq.8)

In Equation 8, T is the number of segments into which the whole time course is divided; i.e., the number of subsequences compared. By default we set T=2, which means we will divide the whole time course into 2 parts, and determine whether the first half and second half are similar. Then, S means how many transition types one includes; e.g. if we want to test stationarity on single transitions, then S = 3 (row-transition, column-transition, diagonal-transition). Next, for \(i_0, i_1, i_2, \ldots, i_r\), \(r\) denotes \(r^{th}\) order transition matrices, \(i_0, i_1, i_2, \ldots, i_r\) indicates how many dimensions we need to do the summation over all types of transitions. Then, such LRX² is asymptotically distributed as Chi-square with degrees of freedom equal to \((T - 1)S^r(S - 1)\). For example, assume we are going to work on 1st order transition matrix with three types of single transitions for T=2—this would tell us whether the frequency of these three first-order transitions is significantly different from the first half of the trial to the second half. In this case, we will sum over \(i_0\), and \(i_1\), with S = 3 (Eq.9). Then,

\[
LRX^2 = 2 \sum_{t=1}^{2} \sum_{i=1}^{S} \sum_{i=1}^{S} n_{i_0, i}(t) \log \frac{p_{i_0, i}(t)}{\hat{p}_{i_0, i}} \sim \chi^2_{S(S-1)(T-1)} \sim \chi^2_{S(S-1)(2-1)}
\]  
(Eq.9)

Meanwhile, \(p_{i_0, i_1, i_2, \ldots, i_r}(t)\) denotes transition probabilities of different time segments and \(\hat{p}_{i_0, i_1, i_2, \ldots, i_r}\) means the associated pooled transition probabilities. In the same example that we showed above (‘BABDCDBCDABABCAB’), we will have 9 transition probabilities and 9 transition frequencies for each time segment (here we have 2 time segments). Then, after we generate the stationarity analytic table (Table. 5), we calculate...
$LRX^2(6) = 2.90, p=.82$. This suggests that the stationarity seems to hold for the sample sequence across the whole time course (the first half is sufficiently similar to the second half).
<table>
<thead>
<tr>
<th>Segment</th>
<th>Ancestral Type</th>
<th>Type II</th>
<th>Type III</th>
<th>Type IV</th>
<th>Type II</th>
<th>Type III</th>
<th>Type IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (first half time course)</td>
<td>Type II</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0.5</td>
<td>0.5</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Type III</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0.5</td>
<td>0</td>
<td>0.5</td>
</tr>
<tr>
<td>BABDCDBC</td>
<td>Type IV</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Type II</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0.67</td>
<td>0</td>
<td>0.33</td>
</tr>
<tr>
<td>2 (second half time course)</td>
<td>Type III</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>DABABCAB</td>
<td>Type IV</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0.5</td>
<td>0.5</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Type II</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>0.5</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>Pooled</td>
<td>Type III</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0.67</td>
<td>0</td>
<td>0.33</td>
</tr>
<tr>
<td>BABDCDBCDAABBCAB</td>
<td>Type IV</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0.67</td>
<td>0.33</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 5. The stationary analysis table for the example sequence 'BABDCDBCDAABBCAB'.

40
Experiment 3

Method

Same experiment as in Experiment 2 (Time-pressure experiment).

Results and Discussion.

In Experiment 2 above, the descriptive statistics (figure 9 & 10) we learned that they could not provide deeper understanding human subjects searching for information under different time pressure conditions. However, when using SED we discovered more inside information that sitting on the transitions between different acquisitions, we compared similarities between different time pressures as well as across two conditions ITP and DTP. Moreover, we still don’t have enough power to tell the underlying mechanisms of information searching strategies, so here we used to further explore such mystery. We used Lag Sequential Analysis (Appendix B) to analyze current TP study. Specifically, we focused on the analyses for determining the order of each trial in different time pressure conditions for every participant. The stacked bar showed order information of different time pressures for ITP (figure 13) and DTP (figure 14). As easily to learn, with increase of time pressure (ITP: from 36s to 6s), participants’ first order information acquisition behaviors dropped dramatically, which indicate under higher time pressure, participants seem lose control of their mental strategies to acquire information. Similarly, for DTP experimental condition, (from 6s to 36s), with lessen of time pressure, first-order behaviors increased, which could be explained due to the gain of mental strategic control during information acquisition experiments. Comparing to standard descriptive statistics, LSA gain more insightful information during information acquisition tracking research paradigms. By indicating the order of internal dynamics, we can gain deeper understanding of human mental strategies in decision making. For example (figure 13), by using SED analysis, we learned that under ITP condition, participants’ behaviors in ‘6s’ have higher difference than in ‘18s’ relative to ‘30s’, then if we switch angle to internal dynamics reflected on order information in figure 13, we could draw the conclusion of such behavioral pattern, it is because that ‘30s’ and ‘18s’ have more first-order behaviors, however, ‘6s’ has barely first-order behaviors. Under the same logic, we could also try to explain DTP in figure 14. The main reason was due to much more zero-order behaviors in all three time pressure. Also, along the analysis of NWD from experiment 2, it seems that when we increase the time pressure (figure 13), people might do a better job under no/low pressure or moderate time pressure (‘36s’, ‘30s’, ‘24s’, ‘18s’), but they may suddenly choke under high pressure, because they didn’t adapt the strategies appropriately. Meanwhile, if we asked participants to face highest time pressure (figure 14) at the beginning (6s), which also, participants have done poor jobs under such time pressure, however, it seems that they could adapt the pressure or suddenly changed their searching strategies while the time pressure lessened. In sum, combining all information above, by using new metrics of NWD and LSA’s order information, we are facing a great opportunity to gain deeper understanding of human mental processes in information acquisition behaviors.
Figure 13. Increasing time pressure data analysis results by using LSA analysis. Y-axis in every sub-figure is all 9 trials for each participant, blue indicates zero order effect (random effect), green indicates first order effect. X-axis for each sub-figure is subject number, current analysis includes 21 participants.
Figure 14. Decreasing time pressure data analysis results by using LSA analysis. Y-axis in every sub-figure is all 9 trials for each participant, red indicates zero order effect (random effect), light blue indicates first order effect. X-axis for each sub-figure is subject number, current analysis includes 28 participants.
General Discussion

In current thesis, we presented two metrics that could have the potential to become the standard daily measurements of human information acquisition strategies and processes. For SED analysis, we learned that such similarity scores could easily incorporate as much information as possible in one sequence, which recorded from process tracing devices. Moreover, such comparison score could be used to differentiate theoretical decision making strategies and empirical ones also could be used under the cases of only empirical similarity comparisons. The two parameters of NWD, substitution matrix and gap, give us more power to explore deeper usage of SED as well, for example, we could use ‘City Block’ instead of ‘Euclidean distance’ to create substitution matrix in order to adjust to the situations which diagonal movements are forbidden. For different gap penalty values we might get different chances to discover different small sub-strategies that underneath one sequence, for example, if one participant used EQW for the first half of experimental time, but LEX for the second one, an appropriate gap value would have a chance to pick up this subtle change which is hardly detected by descriptive statistics (i.e. time per acquisition, number of acquisitions, etc). However, there are two significant limitations of current SED analysis. First, this metric always needs a ‘baseline’ to complete one comparison and then differentiate the similarities between different information acquisition strategies, for example, theoretical sequence in Experiment 1 as ‘baseline’, ‘30s’ and ‘36s’ empirical sequences as ‘baselines’. So, if experimenter wants to use SED analysis just answer the question ‘whether two sequences stand for the same strategy or not’, then one should pick up a third sequence (i.e. control group) as ‘baseline’, then conduct SED analysis and draw the conclusion on the two target sequences. If one experimenter wants to illustrate one sequence should be categorized as specific decision making strategy, then a set of different theoretical sequences should be used as ‘baseline’, then conduct SED analysis between such theoretical ‘baseline’ set and target sequence, finally the comparison with highest similarity score could be the evidence which indicates the strategy that participant used. Second, any similarity score generated by SED analysis doesn’t have absolute meanings if without the help of ‘baseline’, in other words, if NWD is 0.7, we don’t know how significant is such comparison. But we could use non-parametric statistical test to fulfill this shortage. Feusner and Lukoff (2008) applied Monte Carlo Permutation Test on eyetracking data, to test the significance levels of two conditions’ scan patterns, we could adopt the same logic to test the significance levels of information acquisition experiments.

Another metric we presented in this Master Thesis is order information generated by LSA, which reveals internal dynamics during information acquisition processes. If we look back at SED analysis, NWD gave us the power to tell whether two information acquisition sequences are the same or not, LSA go even further, it could tell the reason of such similarity or dissimilarity. For example, in Experiment 3 (same data as Experiment 2), by using LSA, we could explain why participants behave different under different time pressure settings (e.g. high pressure ‘6s’, and low pressure ‘30s’. However, such measurement also has its limitations. First, LSA cannot be used to detect a specific strategy, for example, in Experiment 3, we found several participants’ behaviors are first-order, then we learned their current movements are related to their previous ones,
however, we couldn’t test such relationship is under the hood of LEX or EQW. Second, LSA is not sensitive to the ‘turning point’ or sub-strategies. For example, if one participant used LEX for the first half of one experiment but EQW for the second one. Then order information along could not tell the difference, however, if we could consider other dimensions of LSA, we might get a chance to learn more. For example, stationarity could be used to test the consistency of order information in one trial, and homogeneity could be used to test the consistency of order information across different trials even different blocks or conditions.

In conclusion, the two metrics we presented for capturing the characteristics of information acquisition processes can be used as deeper analytic measurements other than descriptive statistics we used for decades. And such metrics showed high potential to become daily usage and further developments for discovering mental processed that reflected on process tracing data.
Reference


46


Appendix A. String Edit Distance analysis procedure (NWD calculation).
Cristino et al (2010) developed a Matlab toolbox ‘ScanMatch’, which adopted Needleman-Wunsch algorithm as sequence alignment method. We will show how to conduct SED analysis and generate NWD between different conditions.

**String Edit Distance procedure (6s block vs 30s block from ITP)**

*Step 1:* Since we are talking about similarity comparison, we need to select two strings from different blocks first. Because we used similar stimuli for odd and even blocks, so I will choose one trial from 6s, and one trial from 30s as an example.
Table 6. Sequence alignment of two example sequences. Used the substitution matrix above in table 2. All colored cells in the table show the back-tracking path to find out the optimized global alignment to transform ‘CGAGACGT’ into ‘AGACTAGTTAC’. Green colored items mean ‘match’, ‘-’ & grey colored items mean gap insertions, '|' & yellow colored items mean substitutions by using substitution matrix.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>G</th>
<th>A</th>
<th>C</th>
<th>T</th>
<th>A</th>
<th>G</th>
<th>T</th>
<th>T</th>
<th>A</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>-3</td>
<td>-8</td>
<td>-13</td>
<td>-18</td>
<td>-23</td>
<td>-28</td>
<td>-33</td>
<td>-38</td>
<td>-43</td>
<td>-48</td>
<td>-53</td>
</tr>
<tr>
<td>G</td>
<td>-8</td>
<td>4</td>
<td>-1</td>
<td>-6</td>
<td>-11</td>
<td>-16</td>
<td>-21</td>
<td>-26</td>
<td>-31</td>
<td>-36</td>
<td>-41</td>
</tr>
<tr>
<td>A</td>
<td>-13</td>
<td>-1</td>
<td>14</td>
<td>9</td>
<td>4</td>
<td>-1</td>
<td>-6</td>
<td>-11</td>
<td>-16</td>
<td>-21</td>
<td>-26</td>
</tr>
<tr>
<td>G</td>
<td>-18</td>
<td>-6</td>
<td>9</td>
<td>4</td>
<td>-1</td>
<td>-6</td>
<td>6</td>
<td>1</td>
<td>-4</td>
<td>-9</td>
<td>-14</td>
</tr>
<tr>
<td>A</td>
<td>-23</td>
<td>-11</td>
<td>4</td>
<td>-1</td>
<td>-6</td>
<td>-11</td>
<td>1</td>
<td>2</td>
<td>-3</td>
<td>-8</td>
<td>-13</td>
</tr>
<tr>
<td>C</td>
<td>-28</td>
<td>-16</td>
<td>-1</td>
<td>-6</td>
<td>-11</td>
<td>-16</td>
<td>-4</td>
<td>-3</td>
<td>2</td>
<td>-3</td>
<td>-8</td>
</tr>
<tr>
<td>G</td>
<td>-33</td>
<td>-21</td>
<td>-6</td>
<td>-11</td>
<td>-16</td>
<td>-21</td>
<td>-9</td>
<td>-8</td>
<td>-3</td>
<td>1</td>
<td>-4</td>
</tr>
<tr>
<td>T</td>
<td>-38</td>
<td>-26</td>
<td>-11</td>
<td>-16</td>
<td>-21</td>
<td>-26</td>
<td>-14</td>
<td>-13</td>
<td>-8</td>
<td>-4</td>
<td>1</td>
</tr>
</tbody>
</table>

Sequence Alignment: AGACTAGTTAC


CGA ---- GACGT
Step 2: Set up parameters for NWA. Use ScanMatch toolbox to generate substitution matrix by using Euclidean distance. Based on the suggestion by Cristino et al (2010), set gap penalty as 0.

Step 3: Use NWA and dynamic programming technique to finish sequence alignment between two chosen eye movement sequences, and generate one score from [0,1], which the higher score the higher similarity between two sequences.

Step 4: repeat step 1 to 3, to generate an average score between trial 1 under 6s ITP and all 9 trials under 30s ITP.

Step 5: repeat step 4, generate 9 average scores that compared trial 1 to 9 from 6s of ITP and all 9 trials under 30s ITP. And take the average of these 9 scores to get one single score that describing similarity between block 6s and 30s of ITP.

Future direction: Significance level of NWD

While NWD could successfully provide the optimal sequence alignment of a pair of information acquisition sequences—i.e., similarity—it doesn’t have the potential to indicate the significance of such a score. Furthermore, it is also difficult to use conventional parametric statistics to test such significance levels, for example, t-test, Chi-square ($\chi^2$) test. However, we can use Monte Carlo Random Permutation Test (MCRPT; Feusner & Lukoff, 2008), which is a non-parametric statistical test, to test the significance levels of NWD between two big categories of sequences (e.g. see Experiment 1). For a statistical test, it is still very important to generate $p$-values when we want to compare NWD between two conditions. To accomplish such goal, first, we will establish a baseline distance ($d^*$), which indicates the difference between within-condition NWD ($d_{\text{within}}$) and between-condition NWD ($d_{\text{between}}$) (Eq.4):

$$d^* = d_{\text{within}} - d_{\text{between}}$$  (Eq.4)

For example, suppose we have M trials for condition 1 and N trials for condition 2, and then we will compute NWD for all within-condition permutations ( (M-1)! Combinations for condition 1 and (N-1)! Combinations for condition 2), and then take the average of both NWDs to have $d_{\text{within}}$. Next, we will need to compute all permutations of NWD between sequences in condition 1 and 2 ( M*N combinations in total), and take the average to generate $d_{\text{between}}$. The logic behind MCRPT, then, is to put all (M+N) sequences into one pool, then randomly pick M sequences and label as ‘first condition,’ and the remaining N sequences as ‘second condition.’ then recalculate $d^*$. Since for a random grouping, we cannot expect that similarity of two sequences across conditions are significantly higher than that within-condition, then $d^*$ should be very close to zero, on average. Based on the [0, 1] range of NWD provided earlier, a positive $d^*$ statistic denotes that similarities of within-condition sequences are on average further apart than similarities of between-condition sequences. To conduct a permutation test, we need to compare our randomly generated $d^*_{\text{random}}$ with $d^*_{\text{exp}}$ that is derived from the experimental conditions. The null hypothesis for this permutation test is that each grouping is equally likely, and $p$-value is thus the proportion of all $d^*_{\text{random}}$ which are higher than the experimental $d^*_{\text{exp}}$.

By using MCRPT, which is a nonparametric statistical test, we can test the significance levels of different experimental conditions without knowing the underlying similarity distribution, or assuming the NWD of different conditions follow normal distributions.
Appendix B. Lag Sequential Analysis procedure

**LSA Procedure. (For one trial data)**

**Step 1:** Divide eye movement sequence into overlapping digrams by using moving window technique. Then create first order transition matrix from such digram sequence. (Find out the transition frequency that indicate two adjacent eye movements)

Sample sequence: BABDCDBCDABABCAB

<table>
<thead>
<tr>
<th></th>
<th>Attribute 1</th>
<th>Attribute 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Choice A</strong></td>
<td>(A) +</td>
<td>(B) -</td>
</tr>
<tr>
<td><strong>Choice B</strong></td>
<td>(C) -</td>
<td>(D) +</td>
</tr>
</tbody>
</table>

*Figure 15. A fraction of one eye movement sequence, which indicate that subject looked over choice A across all attributes and associated first order transition matrix. Subscript 0 means the first cell in one digram. Subscript 1 means the second move in the same digram.*
<table>
<thead>
<tr>
<th>Step</th>
<th>[t-1, t]</th>
<th>The Moving Time Window</th>
<th>Transition Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-</td>
<td>-</td>
<td>A₀</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>[1]</td>
<td>[B]ABCDDBCAB</td>
<td>A₀</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>[1,2]</td>
<td>[BA]BCDDBCAB</td>
<td>A₀</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>[2,3]</td>
<td>B[AB]DCDBCAB</td>
<td>A₀</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>[3,4]</td>
<td>BA[BD]CDBCAB</td>
<td>A₀</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>[4,5]</td>
<td>BAB[DC]DBCDAB</td>
<td>A₀</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>[5,6]</td>
<td>BABD[CD]BCDAB</td>
<td>A₀</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>[6,7]</td>
<td>BABDC[DB]CDAB</td>
<td>A₀</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>[7,8]</td>
<td>BABDCD[BC]DABAB</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>A₀</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B₀</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>C₀</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>D₀</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>9</th>
<th>[8,9]</th>
<th>BABDCDB[CD]BABABC</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A₀</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>B₀</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>C₀</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>D₀</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>[15,16]</th>
<th>BABDCDBCDABABC[AB]</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A₀</td>
<td>0</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>B₀</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>C₀</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>D₀</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
Step 2: Since we will use Likelihood Ratio Chi-square (LRX) to test the significance of transition matrix, so we need to have the relative expected transition matrix. Also, because such transition matrices are often sparse, so we cannot use normal ways to generate expected transition matrix. Luckily, we can use Iterative Proportional Fitting (IPF), which was created by Fienberg (1970), to generate expected transition frequency matrix for different models. (Bakeman & Quera, 1995).

Step 3: use LRX (equation 1) to calculate significance level between models (equation 2). Degree of freedom = (Row-1)*(Col -1) – (constructive zeros). (more detail see Bakeman & Quera, 1995).

\[ G^2 = 2 \sum (OBS_i) \frac{(OBS_i)}{(EXP_i)} \]  
(Eq.4)

First order models (Lag = 1)
1. Saturated model: \([t-1,t] \) model
2. Independent model: \([t-1][t] \) model.

\[ \Delta G^2_{(t-1)} = G^2_{[t-1][t]} - G^2_{[t-1]} \]  
(Eq.6)

Step 4: If significant difference will be found by testing first order models, then repeat step 1 to 3 again and again until no significant difference will be found. A slight different between higher order (larger than 2) model testing and first order model testing is that first order models only contains two options, but higher order model testing contains 3 options.

Second Order (Lag = 2)
1. Saturated model: \([t-2,t-1,t] \) model
2. Reduced model 1: \([t-2,t-1][t-1,t][t-2,t] \) homogeneous association model
3. Reduced Model 2: \([t-2,t-1][t-1,t] \) conditional association model

Hierarchical Model Testing
\[ \Delta G^2_{(t-2,t-1)} = G^2_{[t-2,t-1][t-1,t][t-2,t]} - G^2_{[t-2,t-1]} \]
\[ \Delta G^2_{(t-2,t)} = G^2_{[t-2,t-1][t-1,t][t-2,t]} - G^2_{[t-2,t-1][t-1,t][t-2,t]} \]  
(Eq.7)