

COMPUTING THE SIMILARITY OF SEQUENTIAL BEHAVIOR

Christopher W. Myers
 Rensselaer Polytechnic Institute
 Troy, NY

Current technology provides researchers' the capability to collect high-density/high-definition data. However, the potential of such capabilities is diminished without the availability of objective analyses. For example, techniques to objectively compare two complete behavioral routines, two subsections within the same routine, or two subsections between two different routines have been elusive. The capability to objectively compare interactive routines of behavior will enable researchers to study the adoption and evolution of such routines. In this paper a technique is proposed to objectively compare behavioral routines, whether the data are obtained from a human or embodied computational model. This technique offers the promise of solving what Anderson (2002) regarded as the non-determinism problem of modeling behavior at the 100-ms level of behavior. The technique is housed within a software tool for integrating and analyzing fixed-location and movement data collected from eyes and cursors, simultaneously (Myers & Schoelles, in press).

INTRODUCTION

There is little disagreement that human-environment interactions require sequential behavior. Yet at the same time there is little consensus on how to analyze such behavioral sequences. When constrained to human-computer interactions, behavioral sequences generally comprise different orderings of *primitive actions* (Anderson, 2002) such as eye movements, eye dwells, mouse movements, mouse dwells, and mouse-clicks. The capability of objectively computing the similarity between behavioral sequences promises to provide insights into interactive routines of behavior.

Searching for a particular software function ('Save As...') within pull-down menus in software applications like Microsoft Word[®] can be used to provide an example of a behavioral sequence. An expert user (A) might first visually dwell the desired menu name ('File'), click on the menu name, visually dwell the desired function within the pull-down menu ('Save As...'), and finally move the mouse and click on the desired function. At the bare minimum, such a behavioral sequence would include a total of four primitive actions: 2 visual dwells and 2 mouse-clicks (excluding eye and mouse movements for simplicity). However, a newcomer (B) to Microsoft products might combine 10 visual dwells with 10 mouse-clicks before completing the same goal, forming a behavioral sequence of 20 components.

In the above example, both users obtained the same goal (i.e., finding and clicking 'Save As...' in the 'File' pull-down menu), and the claims '*user A, the expert, was more efficient and completed the task differently than user B, the novice*' would not be controversial. On the other hand, it is currently impossible to objectively determine *how* different user A and B's behavioral sequences were. This paper proposes a technique to objectively determine the similarity between behavioral routines, whether the data are obtained from a human or embodied computational model. The technique is housed within ProtoMatch, a software tool for integrating and analyzing fixed-location and movement data collected from eyes and cursors, simultaneously (Myers &

Schoelles, in press). The following section briefly describes ProtoMatch, and is followed by an overview of a technique proposed to test the significance of similarities between behavioral sequences.

PROTOMATCH

The increased use of *high-density* data collection (eye and cursor locations sampled at a 60+ Hz rate) within real-world task environments and scaled worlds (Gray, 2002) has accelerated the need to objectively analyze sequential behavior. ProtoMatch is a software tool for analyzing and exploring such sequential data. The tool provides a means to incorporate eye gaze positions and saccades, mouse locations and movements, and mouse-clicks into a unified stream of sequential data derived from log files produced by users behaving in task environments.

ProtoMatch was developed to analyze high-density sequential protocol data throughout a given task from two sources: eyes and mouse-devices (see Table 1). Collecting high-density, sequential protocol data facilitates analyses of behavior occurring at the level of primitive actions (e.g., eye and mouse movements, fixations, dwells, and mouse clicks) (Anderson, 2002; Card, Moran, & Newell, 1983). Sequences of primitive actions provide atomic criteria for cognitive modelers of all types (GOMS, ACT-R, Soar, Epic, etc.) to compare their models.

Table 1. The data source X data type matrix
 Data type

		Movement	Fixed-location
Data source	Eye	scanpaths & saccades	fixations or dwells
	Mouse	mousepaths & movements	fixations, dwells, or clicks

Mouse protocols can be divided into two types: fixed location and movement (see Table 1). Mouse movement protocols are defined as the path a mouse moves over a period of time, or a *mousepath*. Such information can, for example, be used to capture mouse gestures (Moyle & Cockburn, 2003).

Fixed location mouse protocols can be divided into three sub-types: *mouse-clicks* within an interface, the steady positioning of the mouse (*fixations*), and *dwells* (multiple consecutive mouse fixations at approximately the same location). We have found no mention of mouse fixations or dwells in the literature; however, mouse clicks have been regularly used to analyze behavior in task environments (e.g., Card et al., 1983; Fu, 2001; Todd & Benbasat, 2000).

Fixed location eye gaze protocols can be characterized as *fixations* (stable positioning) or *dwells* (multiple fixations at approximately the same location). Eye movement protocols can be characterized as *saccades* (high velocity movements between fixations/dwells) or *scanpaths*, which are constructed through temporally linking fixations and/or dwells (see Table 1).

ProtoMatch was developed, in part, to overcome the traditional separation of fixed location eye gaze and mouse protocols by integrating them. Integration occurs by sequentially interleaving scanpaths and mousepaths into a new protocol

category called *protocol paths*. Protocol paths provide a high-density, high-definition record of behavior, and promise to facilitate a more robust understanding of sequential behavior.

To briefly sum up, fixed location mouse and eye gaze protocols have been traditionally analyzed separately. ProtoMatch overcomes this separation with protocol paths. In order to make good use of behavioral paths (e.g., scanpaths, mousepaths, and the two interleaved as protocol paths), automated analyses need further development. One example is the ability to easily evaluate the similarity between protocol paths. ProtoMatch also houses a score of filters, automated basic analyses, data segmentation, and sequence alignment utilities. Due to limited space, only ProtoMatch's sequence alignment utility will be discussed in detail (see Myers & Schoelles, in press, for a comprehensive description of ProtoMatch).

SEQUENCE ALIGNMENT

The development of software tools for determining the similarity between protocol paths can benefit from a family of algorithms used in bioinformatics. Such algorithms focus on finding subsequences of information within a larger sequence of

Table 2. An example MED matrix produced from the 'immediate' group's data. Each cell contains the minimum edits required to change one sequence (S) into another. The highlighted row represents the prototypical protocol path (PPP) for the group.

S	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	Mean MED:
1	0	8	12	5	9	9	5	6	7	5	6	5	15	5	7	7	7	6	6.89
2	8	0	6	6	7	5	6	6	6	7	4	7	9	5	7	13	13	12	7.06
3	12	6	0	9	7	5	11	12	8	13	8	13	6	9	10	19	19	18	10.28
4	5	6	9	0	8	6	2	5	6	6	5	5	12	4	6	10	10	9	6.33
5	9	7	7	8	0	8	10	9	8	10	7	10	11	7	9	16	15	15	9.22
6	9	5	5	6	8	0	8	9	6	8	3	8	6	6	9	14	14	13	7.61
7	5	6	11	2	10	8	0	5	8	4	5	4	14	4	6	8	8	7	6.39
8	6	6	12	5	9	9	5	0	8	5	6	3	15	5	7	7	7	7	6.78
9	7	6	8	6	8	6	8	8	0	7	5	7	8	6	7	12	12	11	7.33
10	5	7	13	6	10	8	4	5	7	0	5	2	14	4	7	6	6	5	6.33
11	6	4	8	5	7	3	5	6	5	5	0	5	9	3	7	11	11	10	6.11
12	5	7	13	5	10	8	4	3	7	2	5	0	14	4	7	6	6	5	6.17
13	15	9	6	12	11	6	14	15	8	14	9	14	0	12	12	20	20	19	12.00
14	5	5	9	4	7	6	4	5	6	4	3	4	12	0	8	10	10	9	6.17
15	7	7	10	6	9	9	6	7	7	7	7	7	12	8	0	10	9	10	7.67
16	7	13	19	10	16	14	8	7	12	6	11	6	20	10	10	0	1	1	9.50
17	7	13	19	10	15	14	8	7	12	6	11	6	20	10	9	1	0	2	9.44
18	6	12	18	9	15	13	7	7	11	5	10	5	19	9	10	1	2	0	8.83

information through a process called *alignment* (see Thompson, Higgins, & Gibson, 1994, for an example). Sequence alignment algorithms that are similar to those used in the research field of bioinformatics can be applied to determine the similarity between protocol paths.

The alignment algorithm works like a spellchecker, determining the *minimum edit distance* (MED) between two behavioral sequences obtained from humans or embodied computational models. An MED is determined by finding the minimum number of insertions, deletions, and replacements necessary to change one sequence into another. For example, the algorithm’s solution to changing ‘FIREMAN’ into ‘POLICEMAN’ would 1) insert a ‘P’ to the left of the ‘F’; 2) insert ‘O’ to the left of the ‘F’; 3) replace ‘F’ with ‘L’, and 4) replace ‘R’ with ‘C’. Therefore, the MED for changing ‘FIREMAN’ into ‘POLICEMAN’ is 4. Thus, MED provides a measure of similarity – the lower the MED the greater the similarity between two sequences. Changing ‘SCHOOLTEACHER’ into ‘POLICEMAN’ would have an MED greater than 4, and it would be appropriate to conclude that the spelling of ‘FIREMAN’ is more similar to ‘POLICEMAN’ than the spelling of ‘SCHOOLTEACHER’ is to ‘POLICEMAN.’ Replacing the letters in the spelling example with eye and mouse data and their dwelled or clicked GUI items (i.e., ‘F’ = [eye-dwell-1 item1]; ‘I’ = [eye-dwell-2 item2]; ‘R’ = [mouse-dwell-1 item2]; etc.) from a protocol path provides an objective method for determining the similarity between behavioral sequences.

A simple, abstract example will be used to explicate the process of determining statistical differences between experimental groups’ behavioral sequences. Assume there are two between-subjects groups performing a task that is identical in every way except the time it takes to acquire information necessary to complete the goal. The two groups are ‘immediate’ and ‘delayed,’ with 18 participants per group. Although the two groups are accomplishing the same goal, it is important to determine if the groups are using a similar interaction strategy (behavioral sequence) to accomplish the goal.

To determine the statistical significance between behavioral sequences used by the two groups, a *prototypical protocol path* (PPP) representing the ‘average’ sequence of primitive actions must first be found for each group. ProtoMatch produces an MED matrix of sequence comparisons for either between- or within-subject conditions. Table 2 presents an MED matrix where each behavioral sequence, *S*, represents an instance of a single participant’s behavioral sequence, and is compared to all other behavioral sequences for each of the 18 participants from the ‘immediate’ group. Each cell in the matrix is the MED associated with two compared sequences (e.g. *S*₂ & *S*₁ = 8). The matrix was used to determine the PPP, which is simply the sequence with the lowest average MED compared to all other sequences (sequence 11 in Table 2), signifying it was more similar to all other sequences than any other sequence. Once a PPP has been determined for the ‘immediate’ group (*PPP_i*), that same behavioral sequence is then compared to all behavioral sequences from the ‘delayed’ group, producing another MED matrix, similar to Table 2. The MEDs associated with *PPP_i* from the ‘immediate’ and ‘delayed’ group alignments are then

inserted into a one-way analysis of variance. This analysis will determine if the MED variance between *PPP_i* and the group it was derived from (‘immediate’) is equivalent to the MED variance between *PPP_i* and the ‘delayed’ group, and will be interpreted in the same way as all other one-way ANOVAs.

To highlight the differences of the PPPs, prototypes were determined for the ‘immediate’ (*PPP_i*) and ‘delayed’ (*PPP_d*) groups and are presented side-by-side (see Table 3). Both behavioral sequences begin the same; yet on the sixth protocol, the behavioral sequences diverge. That is, the groups are executing the same basic primitive actions (e.g., eye fixations, mouse-clicks, etc.) on the same graphical user interface elements, only in different orders.

Table 3. A comparison of two prototypical protocol paths (PPP) from two groups. ToA, Radar, Target, & Choose represent GUI items necessary to accomplish the task goal. Shaded row represents point of PPP divergence.

Prototypical Protocol Paths		
Protocol number:	Immediate Group (<i>PPP_i</i>)	Delayed Group (<i>PPP_d</i>)
1	Eye-data ToA	Eye-data ToA
2	Eye-data Radar	Eye-data Radar
3	Click Target	Click Target
4	Eye-data ToA	Eye-data ToA
5	Eye-data Radar	Eye-data Radar
6	Eye-data ToA	Click Target
7	Eye-data Radar	Eye-data ToA
8	Eye-data ToA	Eye-data Radar
9	Eye-data Radar	Click Target
10	Click Target	Eye-data Radar
11	Eye-data ToA	Eye-data ToA
12	Eye-data Radar	Eye-data Radar
13	Click Target	Click Target
14	Eye-data ToA	Eye-data ToA
15	Eye-data Radar	Eye-data Radar
16	Eye-data ToA	Click Target
17	Eye-data Radar	Eye-data ToA
18	Click Target	Click ToA
19	Eye-data ToA	Eye-data Choose
20	Eye-data Radar	Click Choose
21	Eye-data ToA	
22	Eye-data Radar	
23	Eye-data ToA	
24	Click ToA	
25	Eye-data ToA	
26	Eye-data Choose	
27	Click Choose	

DISCUSSION

The statistical comparison of behavioral sequences promises to solve what has been dubbed ‘the non-determinism of behavior’ at the 100 ms level of behavior (Anderson, 2002).

That is, Anderson suggests that behavior occurring at the 100 ms level of behavior varies too widely to be accurately modeled within embodied computational architectures such as ACT-R, EPIC, or GOMS. Capturing behavioral sequences using software such as ProtoMatch provides researchers with the unique opportunity to determine how widely behavior varies at such a low level of analysis. The comparisons based on output from sequence alignment algorithms provide a means of determining if groups of participants execute significantly different behavioral routines based on experimental conditions.

Behavioral sequence alignment can be used to compare human data exclusively, human data to model data, or model data exclusively. Human-model comparisons provide a stringent approach to testing the fit of model data to human data, providing atomic criteria for cognitive modelers. For instance, a model may adequately capture human response times or error rates, while inadequately capturing human sequential behavior. In situations where errors and response times are high for humans and models, the models may provide inaccurate insight as to what is driving increased errors or response times.

Finally, the proposed technique does suffer from some insufficiencies. For instance, finding PPPs provides the *mean* strategy from the group. As a result, the technique fails to capture multiple prototypical strategies. It is likely that there is more than one prototypical strategy other sequences would cluster around. Applying procedures such as factor analysis or multidimensional scaling might be a way to capture multiple prototypes within a group.

ACKNOWLEDGMENTS

This work was supported by Office of Naval Research grant #N00014-03-1-0046, and Air Force Office of Scientific Research AFOSR #F49620-03-1-0143, to Wayne D. Gray.

REFERENCES

- Anderson, J. R. (2002). Spanning seven orders of magnitude: a challenge for cognitive modeling. *Cognitive Science*, 26, 85-112.
- Card, S. K., Moran, T. P., & Newell, A. (1983). *The Psychology of Human-Computer Interaction*. Hillsdale, New Jersey: Erlbaum.
- Fu, W.-t. (2001). ACT PRO action protocol analyzer: A tool for analyzing discrete action protocols. *Behavior Research Methods, Instruments, & Computers*, 33(2), 149-158.
- Gray, W. D. (2002). Simulated task environments: The role of high-fidelity simulations, scaled worlds, synthetic environments, and microworlds in basic and applied cognitive research. *Cognitive Science Quarterly*, 2(2), 205-227.
- Moyle, M., & Cockburn, A. (2003). *The Design and Evaluation of a Flick Gesture for 'Back' and 'Forward' in Web Browsers*. Paper presented at the Fourth Australian User Interface Conference on User Interfaces, Adelaide, Australia.
- Myers, C. W., & Schoelles, M. J. (in press). ProtoMatch: A tool for analyzing high-density, sequential eye gaze and cursor protocols. *Behavioral Research Methods*.
- Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 11(22), 4673-4680.
- Todd, P., & Benbasat, I. (2000). Inducing compensatory information processing through decision aids that facilitate effort reduction: An experimental assessment. *Journal of Behavioral Decision Making*, 13, 91-106.