Inferring attention through cursor trajectories

Kiran N. Kumar (knkumar@indiana.edu)
Samuel M. Harding (hardinsm@indiana.edu)
Richard M. Shiffrin (shiffrin@indiana.edu)
Department of Psychological & Brain Sciences, Indiana University, Bloomington

Abstract

The present research infers aspects of spatial attention from movement to targets (and preferably not to foils) of a mouse-controlled cursor on a computer monitor. The long-term goal is a data-rich and rapid assessment technique that can be used to diagnose individual and clinical deficits of attention. The aim of this present research is validating the approach using a college population of subjects. In the experiment, participants attempt to move a cursor toward three spatial positions at which targets appear rapidly but at irregular times, and attempt to inhibit movements toward foils appearing at those positions. We assume that cursor movements toward a position indicates attention has been directed toward that position. A modified Hidden Markov Model (HMM) uses five sources of evidence, all based on parameters to be estimated, to predict the time varying movement of attention: four aspects of cursor movement and a probability that attention will move from one time interval to the next. Five minutes of data are used to estimate parameters for each subject that produce a predicted attention trajectory which best matches what the subject is instructed to do. These parameters are used to predict the attention trajectory for the remainder of the hour of testing. The predictions of attention movements are then matched to the appearance of targets and foils to infer such components of attention as ability to respond to targets vs foils, times to do so, and changes in these components over time. The results illustrate a promising approach to assessment of attention that could likely be employed for both adults and children in clinical settings requiring short testing periods.

Keywords: Attention; Hidden Markov Model(HMM); Individual differences

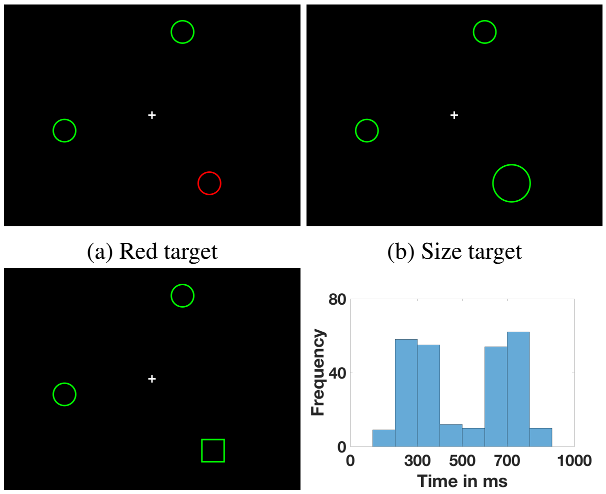
Introduction

Psychologists have been studying various aspects of attention which include ability to focus, maintain, spread attention, and their effects upon behavior. These efforts have accelerated in the last fifty years, in part from articles by Schneider and Shiffrin (1977) and Shiffrin and Schneider (1977), and new methods allowing neural measurements. It would be desirable to have a method that could be used for diagnosis and clinical assessment that could be employed quickly and easily, for example for the testing of children suspected to have various forms of Attention Deficit Hyperactivity Disorder (ADHD). There are many functional aspects of attentional allocation which are measured when studying clinical populations - selective attention, sustained attention, response precision, cognitive flexibility, working memory, temporal information processing, and response inhibition (Mueller, Hong, Shepard, & Moore, 2017). Unfortunately, existing studies of attention require long and extensive periods of testing, and/or equipment that is expensive and cumbersome to utilize. They also require expertise and well trained staff to apply. In this article we demonstrate a methodology that provides a wealth of data about several aspects of attention in a short period of time, with readily and generally available technology, and therefore points towards a promising way to assess and diagnose aspects of attention in individuals. The present research was motivated by an earlier study by Kumar, Chandramouli, and Shiffrin (2015), but that method requires extensive testing for long periods of time and would not have been suitable for assessment and diagnosis. Although our long-range goal is use in clinical and assessment settings, the present study uses college age adults screened to confirm they had not had a diagnosis of attention deficit disorder. The participants are assumed to have a range of attention processes typical of this population. The present research is aimed to provide a proof of concept for the method.

The method is simple in concept: There are three positions displayed on a computer monitor where targets and foils appear at rapid but irregular times. The participant uses a mouse to move a cursor toward targets and attempts to inhibit cursor movements toward foils. The method relies on the fundamental assumption that cursor movements toward a position indicate that attention has been directed toward that position. Cursor movement trajectories provide a wealth of rapidly acquired data, but vary in many and complex ways. As described shortly we therefore use a variant of a Hidden Markov Model (HMM), based on selected aspects of the cursor movements and prior knowledge, to produce a predicted trajectory of spatial attention over time. The predicted attention trajectory can be compared to the sequence and timing of targets and foils to infer the use of attention by each subject tested.

Experiment 1

The three relevant screen positions are indicated by three green circles surrounding fixation at distances of 510 pixels. These positions are fixed throughout testing for all participants. Targets and foils are of three types: A red circle, a larger green circle, and a green square. For each 'block' of testing one of these is designated as a target and the other two as foils. See Figure 1 for examples of target displays. The target choice varies from block to block, for 18 blocks in about one hour of testing. When a target or foil appears, it remains on the screen until the next target or foil appears, with the next one always being in one of the other two screen positions randomly chosen. At that time the previous target or foil reverts to a green circle. When a new stimulus appears it is chosen with probability $\frac{2}{3}$ to be a target and probability $\frac{1}{3}$ to be a foil. If a foil stimulus occurs on a given trial, it is with probability $\frac{1}{2}$ either of two foil choices.



(a) Red target (b) Size target

(c) Square target (d) SOA Distribution

Figure 1: (a,b,c) Displays with salient stimulus at position 1. In a block each of the three stimuli are presented but only one is designated as target.(d) Distribution of times for Stimulus onset asynchrony (SOA) showing a bimodal distribution with mode about 300 and 700ms.

Timing

Targets and foils occur rapidly at irregular times drawn from a bi-modal mixture distribution as shown in Figure 1d. These times are adjusted for each subject on the basis of two practice blocks (see below). For each subject a constant time is added to this distribution to adjust for the speed at which different subjects are able to move the cursor, determined during training. The minimum time added was $128ms$ and the maximum time added was $352ms$, over the 45 subjects. When a stimulus remains on the screen for a relatively long time, it is possible for the cursor to reach and stop at the stimulus; for a short duration the cursor may often be moving toward a given position when the next target (or foil) is perceived at another position. The subjects are instructed to deviate from the previous path and move toward the new stimulus, if it is a target and is perceived before the previous target is reached.

Participants

There were 45 participants, all students of Indiana university, right-handed, with normal or corrected-to-normal vision, and paid for their participation. 5 subjects were excluded from the analysis due to poor performance with no movement in one or more blocks of testing.

Data collection

Each display is presented on a 60-Hz CRT monitor with a resolution of $1600x1200$. Experimental stimuli are generated using matlab image processing toolbox and presented using the psychophysics toolbox extension.(Brainard, 1997). The cursor position was sampled at $1000Hz$, but these positions were downsampled to $100Hz$: i.e. the screen position was collected every 10 ms.

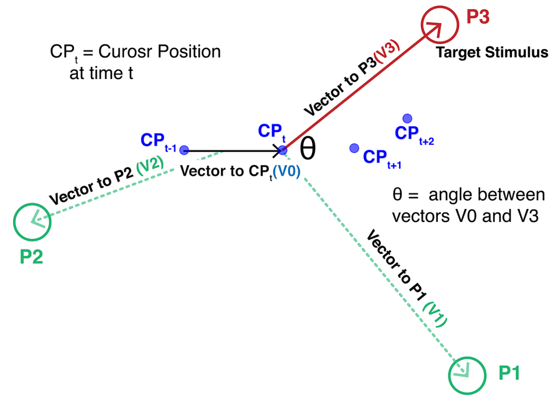


Figure 2: Illustration of trajectory vector (V0) and vector to three display positions (V1,V2,V3) for computing statistics of speed, acceleration, distance, and orientation.

Training and testing

Each participant was given brief practice with two blocks consisting of 45 trials each, but with blue targets and no foils. These trials were used to adjust for each subject the speed at which stimuli appeared. The practice blocks were followed by 18 blocks of testing, with a break of 10 seconds between blocks (a longer break of 30 seconds was placed after 9 blocks of testing). During the break interval, target for the next block was shown at the center of the screen. The 18 experimental blocks consisted of 6 blocks each of targets Red, Size, and Square, in random order. To motivate task performance, participants are notified when they are within 50 pixels from the boundary of a target by changing the cursor to a “hand” symbol. Neither foils nor targets are repeated at the same location on consecutive displays. It can nonetheless be appropriate to keep the cursor on a given position for several displays running, if the next display is a foil, and the display after that is on the position on which the cursor currently resides. A participant sees 270 presentations of targets and foils per block, totaling 4860 trials in about one hour of testing.

Modeling the trajectory of attention

We have no objective measure of the locus of attention at each moment in this task, as is the case for all tasks, but do have good guesses based on trajectories of cursor movements during a trial. For example, if the cursor changes direction and moves rapidly toward a spatial position, we can guess that attention has moved to that position. To assess the placement of spatial attention at each moment we use a modified Hidden Markov Model (HMM): The “hidden” states are the three spatial positions upon which attention likely is placed due to the task requirements. There are multiple sources of evidence that we use to inform the HMM. One is prior knowledge: Lengthy experimentation has shown that attention does not switch very rapidly (Reeves & Sperling, 1986; Vergauwe et al., 2016) and may take a few hundred ms. Thus we wish to employ a constraint implemented by the probability that attention will change state from one moment to the next (one

10 ms interval to the next). If this probability is low then attention will have a tendency to stay in one state. As described later, attention to a given position must be subdivided into two cases, one in which the cursor is moving to that position and the other in which the cursor is stationary at that position. Thus we need two change probability parameters, depending on whether the change is from the moving or stationary state. We do not enforce limits on the size of the change parameters by limiting their size, but rather by use of an objective function when fitting the model's parameters, as described later.

The data used to inform the HMM are aspects of the cursor movements. Cursor movements are complex, noisy and very high dimensional, which leads us to simplify them by using four easy to measure aspects of such movements at each moment in time: Speed, acceleration, and direction toward and distance from the three spatial positions. The way these are calculated at each moment in time is illustrated in Figure 2. The HMM operates at discrete time intervals $10ms$ in duration. At time t we measure the spatial vector from from $t-1$ to t (denoted as $V0$), where speed, s, is determined by the distance moved (vector length), acceleration, a, as the change in speed from $t-2$ to $t-1$ and $t-1$ to t, orientation, o, as the angle between vector $V0$ and the three spatial positions, and distance, d, as the length of vectors from the end of vector $V0$ to the three spatial positions. Each of these measures has a natural way it provides evidence concerning attention placement. For example, consider direction (we use the term orientation): The more directly the current direction of motion is toward a given position, the more likely attention is placed on that position.The amount of evidence (probability) associated with a given set of orientations is unknown, so is parameterized by a plausible two parameter distribution family. Of course orientation is meaningful only when the cursor is moving; when the cursor stops on a position, there is no orientation and this measure provides no evidence. We therefore have two evidence distributions, one when the cursor is moving and another when it is stopped. This idea is implemented by dividing the hidden state for attention at position j into two states, one for movement toward j and another for stopping at position j. The same idea is employed for each of the other cursor movement measures, so that we have six hidden states in the HMM, two for each spatial position.

Associated with these six states are eight probability distributions mapping the four cursor movement measures to the states. Each of these distributions is implemented as a two-parameter family of distributions. Examples of these for particular parameter choices are shown in Figure 4. *Speed* is modeled as a *gamma* with different parameters for moving and stopping states. *Acceleration* is modeled as a *uniform* with different parameters for starting and stopping states. *Distance* is modeled as a *uniform* for the moving state and *gamma* for stopping states. Lastly, *Orientation* is modeled as a *half-normal* for moving and *uniform* for stopping states. The choices of these distribution families were based in part on empirical observations of the cursor movement measures

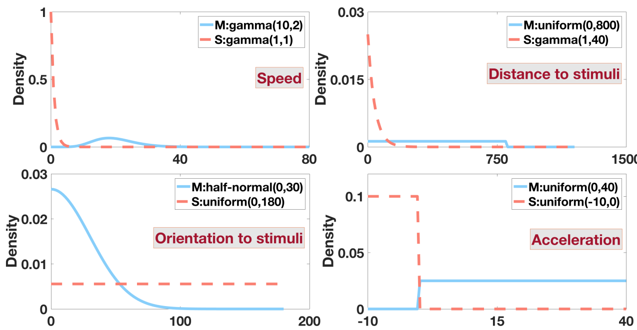


Figure 3: Distributional families of measures - (speed, acceleration, distance, and orientation to stimuli) for moving[M] (blue-solid) and stopping[S] (red-dashed) states

from our study. Later we will show how 'best' parameters are chosen for each subject, including those that specify these eight distributions. As usual, we assume independence of the five sources of evidence, and assume the Markov property: The attention state at time t is determined by the prior probability of attention change, the four measures of cursor movement, and the attention state at time $t-1$, but not on any attention state prior to $t-1$. We use two parameters to describe transitions between states, a high probability for staying in the same state and lower for switching to a state associated with the same spatial position. The remaining probability is divided uniformly for transitioning to other spatial positions. As with any HMM, these assumptions allows us to write the probability of a sequence of attention states in an easy to compute fashion. Conditional on the (unknown) state, Ψ_{t-1}, at time t-1, the state probability at time t, Ψ_t, is a product of the five sources of evidence. Thus given a set of values for the eighteen parameters, Equations 1,2, and 3 give the probability of each set of attention states for any specified experimental duration. Note that it takes time for attention to change and cursors to move, so not all transitions between states are possible. Thus if attention at time t is on position 1, one cannot transit at time t+1 to the resting state at positions 2 or 3; we implement this constraint with zero probability assigned to impossible transitions.

$$p(\Psi_t, \Psi_{t-1}, \dots | s, a, d, o) \tag{1}$$
$$\propto p(s, a, d, o | \Psi_t, \Psi_{t-1}, \dots) p(\Psi_t, \Psi_{t-1}, \dots) \tag{2}$$
$$= p(s|\Psi_t) p(a|\Psi_t) p(d|\Psi_t) p(o|\Psi_t) p(\Psi_t|\Psi_{t-1}) \tag{3}$$

There are an almost uncountable number of sequences of attention states, given each sequence occurs at 10 ms intervals. Although Equation 3 gives the probability of each one, for a given set of parameter values, this is not useful information. Fortunately, standard methods (based on the Viterbi algorithm (S. Godsill, Doucet, & West, 2001)) allows efficient and relatively rapid computational determination of the highest probability sequence of states for a given set of parameter values. This is well known technology and we do not describe it here. Thus the HMM provides us with the maximum probability set of states for each set of parameter values, for a given subject, and for a specified duration of the task (in our implementation, one or two blocks of testing). In reality, we found the system described thus far responded too readily to minor perturbations in cursor movements, so we used a smoothing technique introduced by S. J. Godsill, Doucet, and West (2004) whose window was estimated by impulsive motor feedback reported in Wagner and Smith (2008); this technique smooths across six intervals of 10 ms each (we omit details).

There are many maximum probability sequences of states, one for each set of parameter values. Thus the next step requires selecting the parameters that produce the 'best' sequence of attention states. We do this by finding the parameter values that maximize the objective function defined in Equation 4. This objective function is chosen to reflect what the subject is asked to do in the task: The equation specifies credit allocation when a target occurs in a given position in the period starting 200 ms after the target appears, and ending 200 ms after it is replaced. The first term (T_{target}) gives a unit credit for the first movement of attention to that position. The second term (T_{other}) subtracts one credit for each additional movement to that target position (we do not want too many changes of attention). The third term ($T_{subsequent}$) subtracts one credit for any attention movement to any other spatial position. The last term $f(D_{target})$ is a linear function rewarding early transitions to the target and punishing late transitions to the target (a maximum of two positive credits and a maximum of two negative credits). A similar objective function is used when foils occur, except the first term subtracts rather than adds one credit. We use a standard parameter search algorithm to find the parameter values that maximize this objective function (a simplex algorithm (Nelder & Mead, 1965)).

$$Obj = \arg\max_{\theta}(T_{target} - T_{other} - T_{subsequent} - f(D_{target}) \quad (4)$$

For each subject we found the "best" parameters for the first five minutes of testing (roughly the first two experimental blocks). We then use those parameters to predict attention movements for all subsequent blocks of that subject, and we analyze and show results only for those cross-validation blocks. This procedure is used to reduce over-fitting as far as possible. Using readily available computers, the entire process just described produces a predicted sequence of "best" attention states every 10*ms* for all 18 blocks for each of the 40 subjects in about $240-300$ minutes.

Best parameters

There are 18 parameter values for each of the 40 subjects, and we will not list these. We mention only that the two change probability parameters ranged across subjects from 0.71 to 0.81 for staying in the same state, and from 0.18 to 0.23 for change from a moving state to an allowed stationary state. We did an analysis to see which parameters were particularly sensitive to choice of values. For each parameter we calculated the standard deviation of values across subjects, then produced new predictions for the objective function for values of that parameter plus and minus one standard deviation. The values of the transition parameters made little difference, but the values of the distance parameters in moving states changed the predictions considerably and therefore were more sensitive.

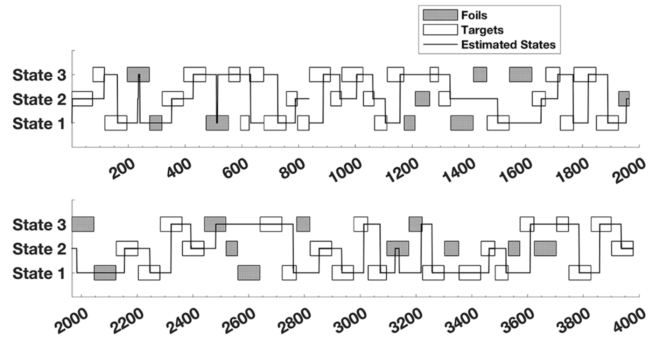


Figure 4: Trajectory of attention showing targets (open rectangles), foils (closed rectangles) and estimated states (solid line) of one subject for 40 seconds duration while searching for a square target.

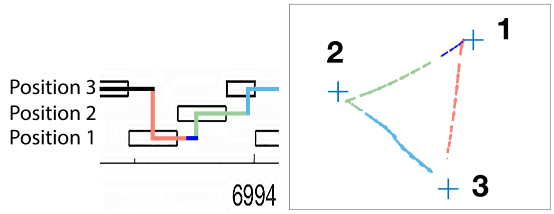
Predicted trajectory of attention

The output of this HMM is a prediction of the sequence of spatial attention states that is suggested by certain cursor movement statistics. These predicted states can be compared to the presented sequence of targets and foils. Figure 4 shows an example for one subject of a short (40 seconds) duration portion in the test blocks following those used to fit parameters. The three spatial positions are shown in the three rows. Target presentations are shown as open boxes, and foil presentations as filled boxes. The width of the rectangles represent the time the stimuli is on the screen. The predictions of the HMM are shown as the solid line superimposed on the presentation sequence. This example shows that the HMM applied to the data for this subject for at least this short time period seems to be operating reasonably: Attention to targets occurs shortly after targets appear, and there are more times attention is given to targets than to foils. Such illustrations are available for all blocks for all 40 subjects, but are far too many to present, so we shall produce various summary statistics to illustrate what the analysis shows about attention changes.

Validation

As mentioned earlier, we have no objective measure of placement of attention. We therefore decided to show the analysis works as desired in two ways:

1) The predicted sequence of attention states can be compared to the raw cursor movements and should operate as expected: E.g. movements toward a position should correspond to a predicted attention state at that position. To illustrate the correspondence in detail we show in Figure 5 in one panel a short snippet of cursor movement (1650 ms.) taken from



(a) Predicted attention transitions by the HMM (b) Actual cursor trajectory observed in the task

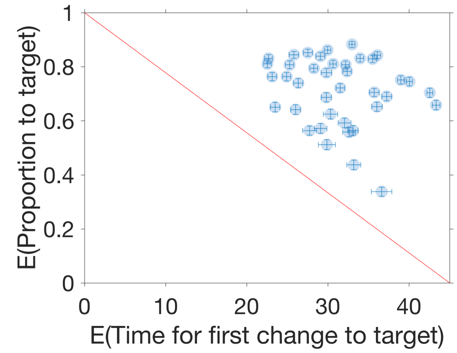
Figure 5: Comparison of predicted attention and cursor trajectory showing a good correspondence. Color of the trajectory in 5b maps to the color of the predicted state in 5b.

one block for one subject. The other panel shows the HMM predictions of the attention states for that same period. One can see that the attention states correspond in a natural way to the cursor movements. This is of course no surprise, because the HMM was built to do exactly this, but examples like this (we checked a large sample of these) serve as a check for programming errors. When the cursor is stationary near one of the target locations (within 200 pixels from center) the HMM predicts attention lies on that location with a median probability of 0.99 across all subjects. If the cursor is moving towards a target with orientation lower than 5 degrees, the model predicts the state with a median probability of 0.93 across all subjects.

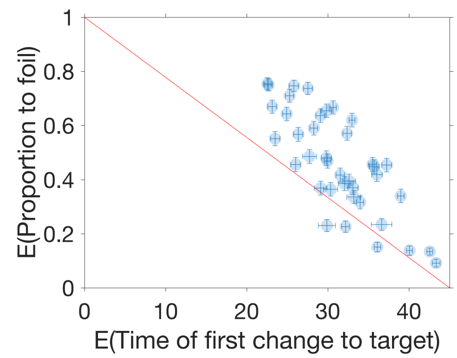
2) The predicted trajectory of attention states should correspond in reasonable ways to the presentation sequence: E.g. There should be more predicted changes of attention to targets than foils, and the times at which these transitions occur should be reasonable based on past studies of the timing of attention shifts. These forms of validation are part of the predictions we produce for each subject, so are placed in the results section.

Results

Figure 6 shows two measures, Figure 6a shows the expected time for first transition and the expected proportion of transitions to targets for each subject. Time to respond to a target does not seem correlated with the probability of doing so, possibly suggesting that times reflect individual differences in reactivity rather than a trade-off. Figure 6b shows the expected time for first transition to target and the expected proportion of transitions for foils: Subjects who take longer to transition to a target have a lower proportion of transitions to foils, suggesting a trade off of speed with reactivity to foils. We note that it could be useful to model the use of and changes of attention by modeling the predictions of the HMM. In Figure 7, the proportion of transitions to targets is compared to those for foils. There are clear individual differences in ability to discriminate the two. All subjects have a higher average proportion of transition to targets than foils although some had great difficulty in discrimination. A paired t-test for each subject shows that all except one with $t(17) = 2.33, p = 0.032$ are able to easily discriminate between targets and foils. The fact that many responses occur to foils suggests that what is being measured is in good part attention to change when a stimulus onset occurs. In further research we plan studies designed to distinguish attention to change from attention to identity.



(a) Time v/s proportion for targets



(b) Time for targets v/s proportion for foils

Figure 6: Mean time for first change to target compared to proportion of changes depicting fast errors.

We note that the large numbers of targets and foils seen by each subject, the enormous amount of cursor movement data collected, and the large number of dynamic predictions of attention states, for each subject, allows us to perform an extremely large number of analyses. We have carried out quite a few of these that space in this short report do not allow us to present. To take just one example, analysis of the cursor trajectories, and of the screen location from which a change of attention state is made, clearly show differences among subjects in the strategies employed to satisfy the task constraints: Some subjects tend to move from one of the three target locations directly to the next; other subjects tend to move continuously to the center of the screen before moving to the next target location. The latter subjects tend to respond quickly to onsets of stimuli, leading us to suspect they might have been experienced video game players, using a strategy that in some sense optimizes both speed and accuracy. Space does not allow us to show these results and a number of others that we have explored.

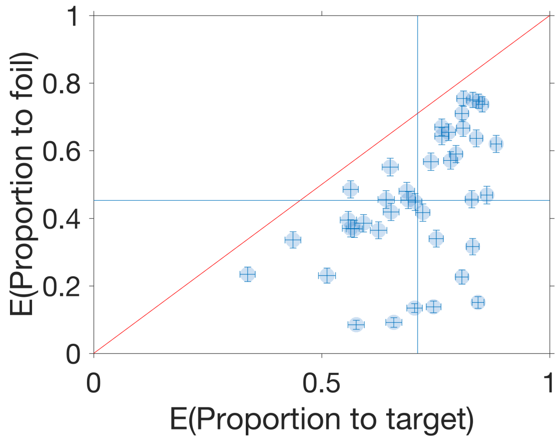


Figure 7: Proportion of responses for targets compared to foils across subjects shows considerable variation across subjects. All subjects show a higher proportion to targets than foils.

Discussion

There are advantages of a procedure to assess attention that operates rapidly and collects large amounts of informative data in short periods of time: the procedures can be used, in principle, for clinical and other forms of assessment of individual differences. There are disadvantages, including the difficulty of reporting many types of analyses and results in a short report. We therefore view this report as a kind of "proof of concept". It would be quite possible to alter the experimental design, aspects of cursor movements fed to HMM, modeling details of the HMM, and procedures for choosing parameters, but retain the benefits of this approach. To take an example of a possibly useful design change, we might want to inhibit strategies by which some subjects move to the screen center before moving to a target. We could use a design in which stimuli appear only briefly enough to be perceived without error, and then revert to the neutral green circles. That would allow the next target to appear at the same location, something that might inhibit movement away from the current target position. We could also employ an error signal if a cursor trajectory passed too close to the screen center. We also note it could prove informative to produce a process model of attention movement that uses the HMM predictions as input data. we note finally that the modeling makes simplifying assumptions, including the assumption that attention at each moment in time is at exactly one spatial location. There are also ways to measure attention with other forms of data such as eye movements (Chuk, Chan, Shimojo, & Hsiao, 2016) or neural measurements. We chose cursor movements because we suspected measurement noise would be less, would not require expensive equipment, careful calibration and expertise in applying them. Computers with monitors are readily available, and subjects who might be assessed with the present procedures are generally familiar with using cursors on computer monitors. The point to emphasize is that we are not wedded to the precise implementation and assumptions made in the present research. The design and procedures nonetheless look very promising as a methodology that could be employed in a variety of forms, and prove useful in practice.

Acknowledgments

The authors would like to thank Robert Nosofsky and David Landy for useful comments and discussion. We thank Daniel Wolpert, Herman van der Kooij and Alfred Schouten for guidance on sensory feedback mechanism for motor control. We also thank the Indiana University Cognitive Science Program for providing research support to KNK and SMH.

References

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial vision*, *10*, 433–436.

Chuk, T., Chan, A. B., Shimojo, S., & Hsiao, J. (2016). Mind reading: Discovering individual preferences from eye movements using switching hidden markov models. In *Proceedings of the 38th annual conference of the cognitive science society, cogsci 2016.*

Godsill, S., Doucet, A., & West, M. (2001). Maximum a posteriori sequence estimation using monte carlo particle filters. *Annals of the Institute of Statistical Mathematics*, *53*(1), 82–96.

Godsill, S. J., Doucet, A., & West, M. (2004). Monte carlo smoothing for nonlinear time series. *Journal of the american statistical association*, *99*(465), 156–168.

Kumar, K. N., Chandramouli, S. H., & Shiffrin, R. M. (2015). Salience, perceptual dimensions, and the diversion of attention. *The American journal of psychology*, *128*(2), 253–265.

Mueller, A., Hong, D. S., Shepard, S., & Moore, T. (2017). Linking adhd to the neural circuitry of attention. *Trends in Cognitive Sciences*.

Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *The computer journal*, *7*(4), 308–313.

Reeves, A., & Sperling, G. (1986). Attention gating in short-term visual memory. *Psychological review*, *93*(2), 180.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. detection, search, and attention. *Psychological review*, *84*(1), 1.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: Ii. perceptual learning, automatic attending and a general theory. *Psychological review*, *84*(2), 127.

Vergauwe, E., Hardman, K. O., Rouder, J. N., Roemer, E., McAllaster, S., & Cowan, N. (2016). Searching for serial refreshing in working memory: Using response times to track the content of the focus of attention over time. *Psychonomic bulletin & review*, *23*(6), 1818–1824.

Wagner, M. J., & Smith, M. A. (2008). Shared internal models for feedforward and feedback control. *Journal of Neuroscience*, *28*(42), 10663–10673.