

# Scan patterns during real-world scene viewing predict individual differences in cognitive capacity

Taylor R. Hayes

Center for Mind and Brain, University of California,  
Davis, CA, USA



John M. Henderson

Center for Mind and Brain and Department of  
Psychology, University of California, Davis, CA, USA



**From the earliest recordings of eye movements during active scene viewing to the present day, researchers have commonly reported individual differences in eye movement scan patterns under constant stimulus and task demands. These findings suggest viewer individual differences may be important for understanding gaze control during scene viewing. However, the relationship between scan patterns and viewer individual differences during scene viewing remains poorly understood because scan patterns are difficult to analyze. The present study uses a powerful technique called Successor Representation Scanpath Analysis (Hayes, Petrov, & Sederberg, 2011, 2015) to quantify the strength of the association between individual differences in scan patterns during real-world scene viewing and individual differences in viewer intelligence, working memory capacity, and speed of processing. The results of this analysis revealed individual differences in scan patterns that explained more than 40% of the variance in viewer intelligence and working memory capacity measures, and more than a third of the variance in speed of processing measures. The theoretical implications of our findings for models of gaze control and avenues for future individual differences research are discussed.**

When viewing a scene, we actively move our eyes three to four times each second to sample the visual environment (Henderson, 2003). The series of eye movements during scene viewing can broadly be categorized into periods where our eyes are held relatively stable (fixations) and periods where our eyes rapidly move between two different spatial locations (saccades). During fixation periods, the area of the eye's retina with the highest acuity is directed to a specific location within the scene and high-quality visual information is acquired. In contrast, limited visual information is acquired during the rapid saccadic eye movements between fixations—a phenomena known as saccadic suppression (Matin, 1974; Thiele,

Henning, Kubischik, & Hoffman, 2002). The result of viewing a real-world scene is a complex, sequential pattern of fixations and saccades called “scanpaths” or “scan patterns” (Noton & Stark, 1971a, 1971b; Stark & Ellis, 1981; see Figure 1). Scan patterns provide a rich description of how overt visual attention is used to sequentially filter our visual environment in an effort to satisfy ongoing perceptual, cognitive, and behavioral goals (Churchland, Ramachandran, & Sejnowski, 1994; Findlay & Gilchrist, 2003), and have the potential to provide important insights into the underlying mechanisms of gaze control during scene viewing.

Pioneering work on scene viewing by Buswell (1935) and Yarbus (1967) produced early qualitative findings that described differences in scan patterns related to the properties of the scene stimulus, the scene task, and the viewer. Since this early scene work, there has been a great deal of research investigating how image properties such as color, intensity, and orientation contribute to bottom-up scene saliency (Itti & Koch, 2000, 2001; Koch & Ullman, 1985; Torralba, 2003), and the ways in which visual saliency influences where people look in a given scene (Bruce & Tsotsos, 2009; Koehler, Guo, Zhang, & Eckstein, 2014; Parkhurst, Law, & Niebur, 2002). More recently, there has been a renewed interest in attempting to quantify the extent to which eye movements reflect top-down cognitive relevance as a function of scene viewing task. For instance, it has been shown that scene task can be predicted at above-chance levels from fixation number, fixation duration, and saccade amplitude distributions (Borji & Itti, 2014; Henderson, Shinkareva, Wang, Luke, & Olejarczyk, 2013; Kardan, Berman, Yourganov, Schmidt, & Henderson, 2015). In addition, it has been shown that target scene objects that are in semantically appropriate, low-salience locations are quickly located and fixated, whereas visually salient, but cognitively irrelevant scene locations are rarely fixated (Henderson, Malcolm, & Schandl, 2009). These findings highlight

Citation: Hayes, T. R., & Henderson, J. M. (2017). Scan patterns during real-world scene viewing predict individual differences in cognitive capacity. *Journal of Vision*, 17(5):23, 1–17, doi:10.1167/17.5.23.

doi: 10.1167/17.5.23

Received October 10, 2016; published May 31, 2017

ISSN 1534-7362 Copyright 2017 The Authors

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.



Downloaded From: <http://jov.arvojournals.org/> on 04/19/2018



Figure 1. Example viewer scan pattern during scene memorization. The viewer was instructed to memorize the scene for a later memory test. During 12 s of viewing the viewer made 42 fixations. The red circles show each fixation location and the red lines indicate saccades between fixations. The white numbers indicate the sequential order of Fixations 1 through 42.

that gaze control is also strongly modulated by top-down task demands during scene viewing. Together these lines of research are both theoretically important as they identify and quantify how bottom-up stimulus properties and top-down cognitive relevance each contribute to gaze control during scene viewing (Koehler et al., 2014; Torralba, Oliva, Castelano, & Henderson, 2006).

One area that has received significantly less attention is the relationship between individual differences in scene scan patterns and viewer individual differences. This is surprising considering the majority of studies that have reported scan patterns during scene viewing have noted that different viewers often produce qualitatively different scan patterns under constant stimulus and task demands (Buswell, 1935; DeAngelus & Pelz, 2009; Henderson & Hollingworth, 1998, 1999; Noton & Stark, 1971a, 1971b; Stark & Ellis, 1981; Underwood, Foulsham, & Humphrey, 2009; Yarus, 1967). The lack of work in this area may be due to the computational difficulties associated with quantifying scan patterns, which often prevents researchers from analyzing them (Hayes et al., 2011). Nevertheless, scan pattern differences when the scene stimulus and task are held constant suggest the need for a better understanding of the role of viewer individual differences. The goal of the present study was to quantify the strength of the association between individual differences in scan patterns and viewer individual differences during scene viewing. Specifically, we investigated three cognitive dimensions of individual difference including

intelligence, working memory capacity, and speed of processing.

Individual differences in scan patterns associated with viewer cognitive capacities were extracted using a powerful technique for scan pattern analysis called Successor Representation Scanpath Analysis (SRSA; Hayes et al., 2011, 2015). SRSA uses temporal difference learning (Sutton, 1988) to capture statistical regularities in scan patterns in a fixed-size matrix called a “successor representation” (SR; Dayan, 1993) that can be aggregated across trials and analyzed with standard multivariate methods. SRSA was used to quantify the strength of the association between individual differences in scan patterns and individual differences in viewer cognitive capacities by identifying individual differences in scan patterns during scene encoding that predicted viewers’ intelligence, speed of processing, and working memory capacity scores assessed via a separate cognitive test battery.

Our results produced clear support for a strong association between individual differences in viewer scan patterns and cognitive capacities. SRSA identified individual differences in scan patterns during scene viewing that explained more than 40% of the variance in intelligence and working memory capacity scores, and more than a third of the variance in speed of processing scores across participants. Moreover, the SRSA results were interpretable in terms of individual differences in global information processing strategies during scene encoding, such as how participants shifted their overt attention from central and peripheral scene information, and regularities in how participants scanned the scenes horizontally and vertically. The implications of these findings are broad as they suggest individual differences in the cognitive capacities of the viewer are strongly associated with how overt attention is deployed to encode complex visual information in scenes.

## Method

### Participants

Seventy-nine University of South Carolina undergraduate students with normal or corrected-to-normal vision participated in the experiment. All participants were naive concerning the purposes of the experiment and provided informed consent.

### Apparatus

Eye movements were recorded with a SR Research EyeLink 1000 plus tower mount eye tracker (spatial

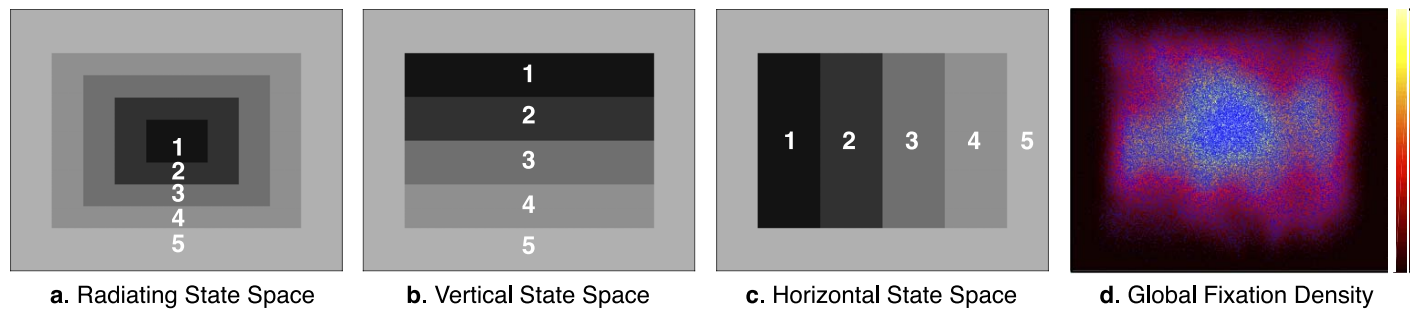


Figure 2. State spaces used to define sequential scan patterns during scene viewing. Scan patterns during scene viewing were defined by mapping fixation positions to three different state spaces. The radiating state space (a) measured viewer tendencies to shift their overt attention between central and peripheral scene information. The vertical and horizontal state spaces (b and c) measured observers' tendencies to shift their overt attention vertically and horizontally. Each of the state spaces contained an outside state 5 that reflected the center bias observed in the global fixation density (d) across all scenes and participants ( $N = 65$ ). Each state space was applied globally across all 40 scenes.

resolution 0.01) sampling at 1000 Hz (SR Research, 2010b). Participants sat 90 cm away from a 21-in. monitor, so that scenes subtended approximately  $33^\circ \times 25^\circ$  of visual angle. Head movements were minimized using a chin and forehead rest. Although viewing was binocular, eye movements were recorded from the right eye. The experiment was controlled with SR Research Experiment Builder software (SR Research, 2010a).

## Scene stimuli and task procedure

Stimuli consisted of 40 digitized photographs of real-world scenes. The real-world scene stimuli included a variety of indoor and outdoor environments. Five of the scenes contained people. Participants were instructed to memorize each scene in preparation for a later memory test that was not administered. Each trial began with a fixation on a cross at the center of the display for 300 ms. Following fixation, each scene was presented for 12 s while eye movements were recorded. Scenes were presented in the same order across all 79 participants. After completing the scene memorization task, subgroups of between 30 and 40 participants completed a series of individual difference measures of intelligence, speed of processing, and/or working memory capacity (see Appendix B for details). The individual difference measures were administered to subgroups of participants due to session time constraints. Eye movements were not recorded during the individual difference test battery.

## State space definitions

Three different state spaces were defined a priori to capture simple scene viewing tendencies and applied to eye movements to produce scan pattern sequences

across each scene (see Figure 2). Each state space spanned the full display ( $1024 \times 768$  pixels) and was used to examine different patterns in how participants shifted their overt attention during scene viewing. The radiating state space consisted of a series of radiating rectangular areas of interest (AOIs) from the scene center to the periphery, and was used to represent observers' tendencies to shift their overt attention between more central and peripheral scene information. The vertical state space consisted of four equal rectangular horizontal AOIs, and was used to represent observers' tendencies to shift their overt attention vertically across each scene. The horizontal state space consisted of four equal rectangular vertical AOIs, and was used to represent observers' tendencies to shift their overt attention horizontally across each scene. Note each of these three state spaces contained an outside border that reflected the central fixation bias, the commonly observed phenomena in which participants concentrate their fixations more centrally and rarely fixate the outside border of a scene (Tatler, 2007). A state number of five was chosen for each state space instead of 10, as was used in previous SRSA applications (Hayes et al., 2011, 2015), because there were too few fixations per scene viewing trial to support a state resolution of 10. The radiating, vertical, and horizontal state spaces were each applied separately to all 40 scenes to map each participant's fixation positions (i.e.,  $x$  and  $y$  coordinates) to one of the five distinct states within each state space.

## Eye movement data

A 13-point calibration procedure was performed at the start of each session to map eye position to screen coordinates. Successful calibration required an average error of less than  $0.49^\circ$  and a maximum error of less



than  $0.99^\circ$ . Fixations and saccades were segmented with EyeLink's standard algorithm using velocity and acceleration thresholds ( $30^\circ/\text{s}$  and  $9500^\circ/\text{s}$ ; SR Research, 2010b). The eye movement data was imported into Matlab using the EDFConverter tool, which converted the EyeLink data file to text that was then imported into Matlab. In Matlab, the eye movement data from each participant was inspected for excessive artifacts caused by blinks or loss of calibration due to incidental movement by examining the mean percent of signal across all trials (Holmqvist et al., 2012; Holmqvist et al., 2015). Fourteen participants with less than 75% signal were removed, leaving 65 participants that were tracked very well (mean signal = 91.74%). Traditional eye movement metrics such as fixation duration, saccade amplitude, and fixation number were computed for each trial. In addition, scan patterns across the three different state spaces (Figure 2) were computed using the  $x$  and  $y$  gaze positions of each fixation within each scene trial. The first fixation of each trial was discarded because it was always at the center of the display as a result of the pretrial fixation period, and thus uninformative.

### Successor representation scanpath analysis

SRSA was used to capture statistical regularities in scan patterns within each of the three different state spaces and predict individual differences in the cognitive capacities scores of the participants. SRSA quantifies regularities in scan patterns using temporal-difference learning (Sutton, 1988) to construct a fixed-size matrix called an SR (Dayan, 1993). The key idea behind SRSA is that upon observing a transition from one state (i.e., a defined AOI within a state space) to another, instead of simply updating the transition probability from the first to the second state, SRSA associates the first state with the second state and all expected subsequent states based on prior visits to the second state. In this way the SRSA algorithm learns to predict future scan patterns based on past scan patterns. After traversing a scan pattern for a given scene, the resulting SR can be conceptualized as having extracted the statistical regularities in temporally extended scan patterns. Specifically, an SR matrix contains, for each state, the temporally discounted number of expected future fixations to all states (Dayan, 1993). Given the uniform size of SRs and a commonly defined set of states, the SR matrices from different observers and/or trials can be analyzed using standard statistical methods to identify significant pattern regularities for various comparisons of interest. SRSA has previously been successfully applied to study individual differences in problem-solving strategies during matrix reasoning (Hayes et al., 2011) and the

role of strategy refinement in pre-post designs using matrix reasoning tests (Hayes et al., 2015). In the present study, SRSA used individual differences in scene viewing scan patterns to predict individual differences in viewer intelligence, speed of processing, and working memory capacity.

As described by Hayes et al. (2011, 2015), the first step in SRSA is to convert the eye movements for each trial into a trial SR. For the sake of simplicity, we will describe the SRSA analysis in terms of a single state space (i.e., the radiating state space), but the exact same procedure was applied to all state spaces shown in Figure 2. For each scene trial, each fixation was mapped to one of the five states in the radiating state space based on the fixation position coordinates ( $x$  and  $y$ ). This converts a series of fixation positions into a scan pattern across the five distinct states within the radiating state space. After mapping eye movement positions to the radiating state space, an SR for each trial can be computed. An SR (Dayan, 1993) was calculated for each trial scan pattern, resulting in one  $5 \times 5$  SR matrix  $M$  per trial for each participant. To calculate the trial SR matrix, each trial SR matrix is initialized with zeros and then updated for each transition in the scan pattern. Consider a transition from state  $i$  to state  $j$ . The  $i$ th column of the matrix—the column corresponding to the “sender” state—is updated according to:

$$\Delta M_i = \alpha(I_j + \gamma M_j - M_i), \quad (1)$$

where  $I$  is the identity matrix, each subscript picks a column in a matrix,  $\alpha$  is a learning-rate parameter ( $0 < \alpha < 1$ ), and  $\gamma$  is a temporal discount factor ( $0 < \gamma < 1$ ). The learning rate parameter  $\alpha$  controls the incremental updating and  $\gamma$  controls the amount of temporal discounting. The  $\gamma$  parameter is the key to extending the event horizon to encompass both immediate and long-range transitions—it includes the discounted future states in the prediction from the current state. For example, suppose a participant scans a scene systematically moving from the center outward twice:  $1 \rightarrow 2 \rightarrow 3 \rightarrow 1 \rightarrow 2 \dots$ . Then the successors of Location 1 will include *both* Location 2 and, weighted by  $\gamma$ , Location 3. Therefore, when  $\gamma$  is set to zero the SR is equivalent to a first-order transition matrix and as  $\gamma$  increases, the event horizon is extended farther and farther into the future. After traversing the whole scan pattern, the estimated trial SR matrix approximates the ideal SR matrix, which contains the temporally discounted number of expected future fixations on all state AOIs (SR matrix rows), given the participant just fixated on any individual state AOIs (SR matrix columns). Note that the entries in the SR matrix are not probabilities; they are (discounted, expected) numbers of visits. Note also that the learning parameter  $\alpha$  does not reflect a cognitive learning rate, but only

the learning rate that optimizes the temporal-difference learning algorithm. At this stage, the data set for the radiating state space in our example, consisted of 40  $5 \times 5$  trial SR matrices per participant, one for each scene viewed.

The second step, how to aggregate SRs, is dependent on the question of interest—in our case we wished to quantify the strength of the association between individual differences in scan patterns during scene viewing and cognitive individual differences in the viewers. Since we were interested in examining individual differences at the participant level, we collapsed across trials by averaging the 40 trial  $5 \times 5$  SR matrices for each participant, resulting in one mean  $5 \times 5$  SR matrix that summarized the scan patterns of the corresponding participant during the scene encoding task within a given state space. Each participant SR matrix was reshaped to a vector of 25 features. To reduce the dimensionality of this 25 feature space and prevent overfitting, we performed a principal-component analysis (PCA; Everitt & Dunn, 2001). PCA is a technique for reducing dimensionality by finding the most informative viewpoints (i.e., variance-maximizing orthogonal rotations) of a high-dimensional space. The result is a set of linear orthogonal variables called principal components. Following standard PCA practice, we rescaled each SR feature so that it had zero mean and unit variance across the participants. Conceptually, these components represent dimensions of individual differences in scan patterns. They are expressed mathematically as orthogonal basis vectors in the 25-dimensional SR space. Across all the individual difference measures tested, the first 20 principal components retained over 98% of the variance in the SR data. Given that the first 20 components retained on average over 98% of the variance (first 10 components: 90%; first 15 components: 95%), each participant was characterized by 20 projections onto this rotated basis, following Hayes et al. (2011, 2015). These 20 component projections served as potential candidates for the five projections ultimately selected by the hierarchical regression in the next processing step.

The final step in SRSA is to optimize and cross-validate the model fit between the SR projections and the current target individual difference measure (i.e., the participant scores on the intelligence, processing speed, and working memory capacity measures). The same two-tier algorithm as Hayes et al. (2011, 2015) was used to maximize the fit. In the inner loop, the algorithm calculated the mean participant SRs for given parameters  $\alpha$  and  $\gamma$  (Equation 1), then calculated the first 20 principal components and the corresponding projections for each participant, picked the five projections that correlated most strongly with the

target individual difference measure, and constructed a linear regression model with these five predictors.

In the outer loop, a Nelder-Mead optimization routine searched for  $\alpha$  and  $\gamma$  that maximized the multiple regression coefficient of the inner-loop model. To guard against overfitting, we performed leave-one-out cross-validation to test the generalization performance of the two-tier fitting algorithm. That is, we partitioned the data into a series of training and test sets where each participant is left out in turn. We ran the two-tier algorithm on the training set. The parameters  $\alpha$  and  $\gamma$  optimized on the training set were then used to calculate the SRs for the scan pattern sequences of the left out participant. Finally, we calculated the model's prediction of the current cognitive capacity measure score by multiplying the left out participant's mean SR matrix by the prediction weight matrix (i.e., the sum of the five best principal components scaled by their respective regression coefficients) from the training set. We repeated this process for each participant. This produced a predicted individual difference score for each left out participant, each one based on a model that had no access to the data that was subsequently used to test it.

For all SRSA analyses a goodness-of-fit  $R^2$  across all participants and a leave-one-out cross-validated  $R^2_{cv}$  fit are reported. The cross-validated fit is a much better estimate of the generalization performance than the goodness-of-fit  $R^2$  (Hastie, Tibshirani, & Friedman, 2009; Haykin, 2009). The goodness-of-fit  $R^2$  is inflated because it reflects not only genuine regularities in the population, which will generalize to new cases, but also the idiosyncrasies of the training sample, which will not. SRSA was systematically performed in this same way for each state space definition (i.e., radiating, vertical, and horizontal) to predict each cognitive capacity measure (i.e., Raven's score, SAT score, Trail A score, Trail B score, operation span, reading span, and general intelligence).

### Procedure for aggregating across SRSA leave-one-out sets

The goodness-of-fit ( $R^2$ ) SRSA models produced a single set of five principal components and one prediction weight matrix (the sum of the five principal components scaled by their respective regression coefficients) across all participants. However, as discussed above, leave-one-out cross validation performance is a superior measure of model generalization performance. The leave-one-out procedure produced five principal components and one prediction weight matrix for each leave-one-out set, resulting in  $N$ -participant sets of components and prediction weights. The cross-validation prediction weight matrix is easily

computed as the mean across the  $N$ -participant prediction weight matrices. However, the principal components require a more sophisticated procedure, because while they are highly consistent across runs their rank order can occasionally shift between different leave-one-out sets. Therefore, components must first be clustered into the appropriate group prior to averaging.  $K$ -means clustering was used to group the principal components into the appropriate five groups, and then the average was computed resulting in five mean principal components for each cross-validated model.

## Procedure for interpreting SRSA weights

Unlike previous successful applications of SRSA (Hayes et al., 2011, 2015), which applied SRSA to a well-defined problem space (i.e., a matrix reasoning task), the real-world scene encoding task is less constrained. As a result, it makes interpreting the prediction weight matrices more difficult. However, one of the major advantages of SRSA is the prediction weight matrices and principal components are interpretable. The main barrier to interpretation in this less constrained scene viewing task is providing a way to distill and visualize the higher order sequential patterns that are being captured—a notoriously difficult visualization task (Aigner, Miksch, Schumann, & Tominiski, 2011). To assist in the interpretation of the higher order sequential patterns captured by the SRSA prediction weight matrices in the less constrained scene encoding task, a general procedure was developed to identify the most illustrative scene scan patterns for each SRSA individual difference model (i.e., the individual scene scan patterns that were the most strongly positively and negatively correlated with the SRSA prediction weights).

A simple procedure was used to search for illustrative example scan patterns for each cognitive individual difference SRSA model. For each individual difference measure the 5 highest and 5 lowest scoring participants were selected and the optimal SRSA parameters ( $\alpha$  and  $\gamma$ ) were used to convert each scene trial scan pattern into a trial SR. The correlation between the trial SRs from the five highest and five lowest scoring participants were then correlated with the mean cross-validation prediction weight matrix from their respective SRSA model. These correlations represented the association strength between the SRSA prediction weights and the trial scan patterns, where positive correlations were indicative of higher cognitive capacity scores and negative correlations were indicative of lower cognitive capacity scores. In order to identify the most illustrative scan patterns, the 40 correlations (one for each scene scan pattern) from each of the five lowest

scoring participants were subtracted from each of the five highest scoring participants, resulting in 25 difference vectors of length 40. The maximum value across all of these correlation differences was selected as the most illustrative positive and negative scan pattern pair for each individual difference SRSA model, within a common scene. The most illustrative example scan patterns are shown in Figure 5 for the SRSA models that are highlighted with asterisks in Table 1, including the model cross-validated prediction weights, the most illustrative trial SRs, and the corresponding scan patterns plotted as a function of change in state and state transition length.

## Results

The 65 participants that completed the scene encoding task and met the eye tracking signal criterion produced a total of 93,485 fixations with an average of 1,438 ( $SD = 167$ ) fixations per participant. The mean participant fixation duration across all scene trials was 283 ms ( $SD = 43.2$  ms). The mean participant saccade amplitude was  $3.5^\circ$  ( $SD = 0.44^\circ$ ) and the mean participant fixation number per scene trial was 35.9 ( $SD = 6.0$ ). Figure 3 shows the mean, standard deviation, and the distribution of participant subgroup scores on each of the collected cognitive individual difference measures including Raven's score (fluid intelligence), SAT score (crystallized intelligence), Trail A and B scores (speed of processing), and operation and reading span (working memory capacity).

## Scene scan patterns and cognitive capacity

Recall our goal is to quantify the strength of the association between individual differences in scan patterns during scene viewing and individual differences in the cognitive capacities of viewers. The goodness-of-fit and cross-validated SRSA prediction performance is shown in Table 1 for each cognitive individual difference measure and state space. The SRSA results showed that individual differences in scan patterns could explain large amounts of variance in the cognitive capacity measures. The radiating state space produced the best overall prediction for the intelligence measures explaining over 40% of the variance in Raven's score ( $R^2_{cv} = 0.43$ ) and SAT score ( $R^2_{cv} = 0.45$ ). The radiating state space also produced the best performance for operation span, explaining more than 40% ( $R^2_{cv} = 0.45$ ). The vertical state space produced the best performance for the speed of processing measures explaining more than a third of the variance in Trail B score ( $R^2_{cv} = 0.34$ ) and more than 40% of the variance



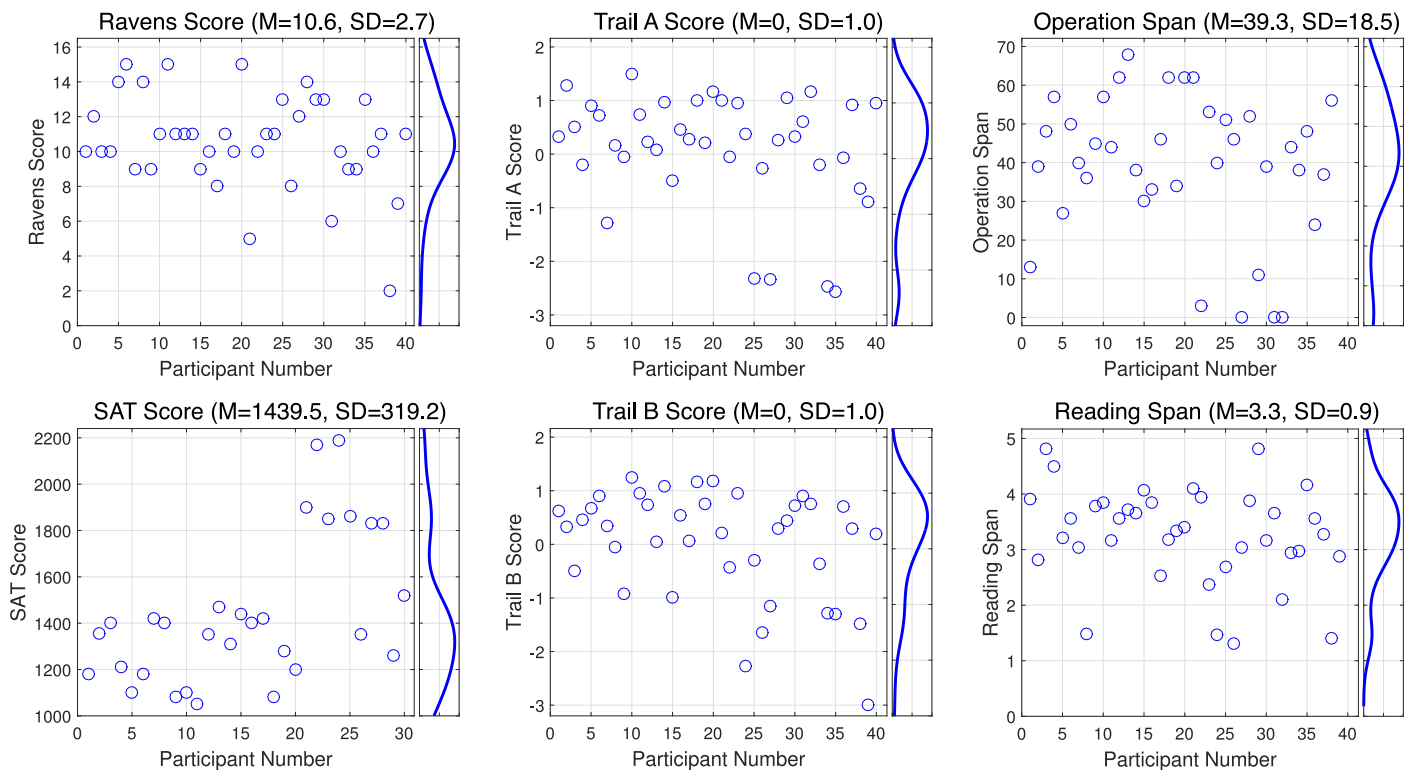


Figure 3. Observed individual difference measure score scatter plots with corresponding probability density histograms.

in Trail A score ( $R^2_{cv} = 0.42$ ). The vertical state space also explained more than 40% of the variance in reading span scores ( $R^2_{cv} = 0.42$ ) and general intelligence scores ( $R^2_{cv} = 0.41$ ), and half the variance in SAT score ( $R^2_{cv} = 0.50$ ). These results provide support for a strong association between individual differences in scan patterns during scene viewing and the underlying individual differences in the cognitive capacities of the viewers.

Looking generally at the SRSA prediction performance in Table 1 reveals that each state space was able to predict multiple individual differences measures well. This suggests that these state spaces were well suited to forming representations of temporally extended scan patterns during scene viewing that are related to underlying individual differences in cognitive capacity. However, it is worth noting that the radiating state space could not predict Trail B score well ( $R^2_{cv} = 0.07$ ), the vertical state space could not predict Raven's score well ( $R^2_{cv} = 0.15$ ), and the horizontal state space had the weakest overall performance of the state spaces that were tested. These weaker predictions suggest some state spaces do not capture the scan pattern regularities that are relevant for certain dimensions of cognitive individual differences. This point is reinforced by the fact that certain state space definitions seemed to be very well suited to predicting specific cognitive capacities. For instance, the radiating state space seemed particularly well suited to capture scan patterns

associated with individual differences in intelligence, while the vertical state space seemed best suited to capture scan patterns related to speed of processing measures. The working memory capacity measures (i.e., operation span and reading span) were split between the radiating and vertical state space.

Figure 4 shows the mean prediction weights and mean principal components for the cross-validated SRSA model for each individual difference measure highlighted with an asterisk in Table 1. In the case of the intelligence and speed of processing measures, where there was a clear preferred common state space, we will focus on the SRSA models from their preferred state space. In the case of the working memory capacity measures, where there was no clear preferred common state space, we will highlight the most predictive ( $R^2_{cv}$ ) SRSA model. Recall the mean cross-validated prediction weights are the sum of the principal components scaled by their respective regression coefficients and provide a summary of the five principal components. For sake of simplicity and brevity, we will focus our interpretation on the SRSA model prediction weights.

## Intelligence

The radiating state space seemed best suited for capturing scan patterns associated with individual differences in intelligence. SRSA using the radiating

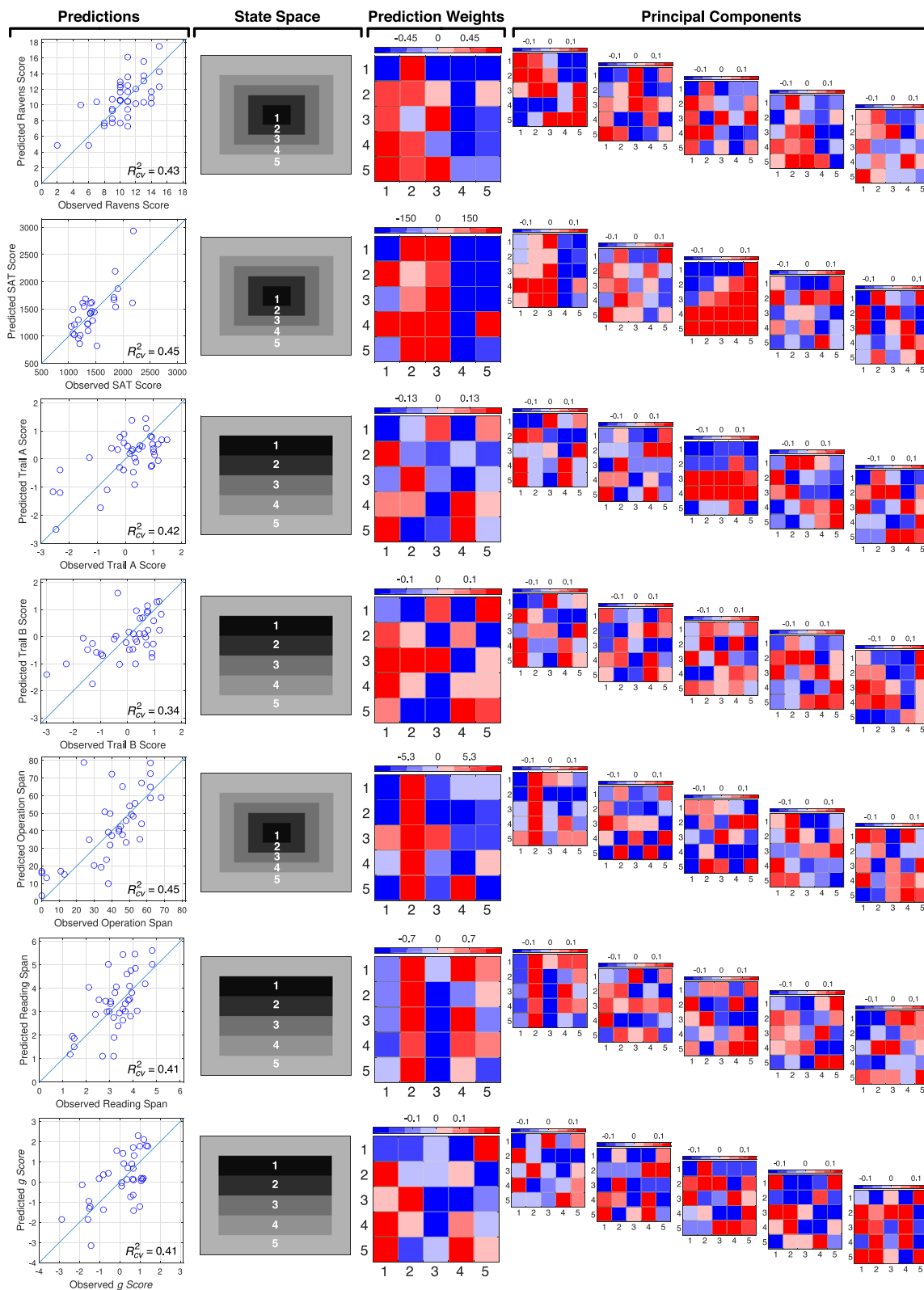


Figure 4. Individual difference score observations and predictions, state space, prediction weights, and principal components for each cognitive capacity SRSA cross-validated model. The Predictions column shows the observed and SRSA predicted cognitive capacity scores and their squared correlation, where the line represents a squared correlation of 1. The State Space column shows the state space definition for each model. The Prediction Weights column shows the mean prediction weights across the leave-one-out fits for each individual difference measure. Finally, the five mean principal components across the leave-one-out fits are shown for each cross-validated SRSA model ranked according to the mean amount of variance they captured across the training sets. Positive values associated with higher individual difference scores are shown in red and negative values associated with lower individual difference scores are shown in blue. In the prediction weights and principal components the x-axis represents the sender state and the y-axis represents the receiver state.



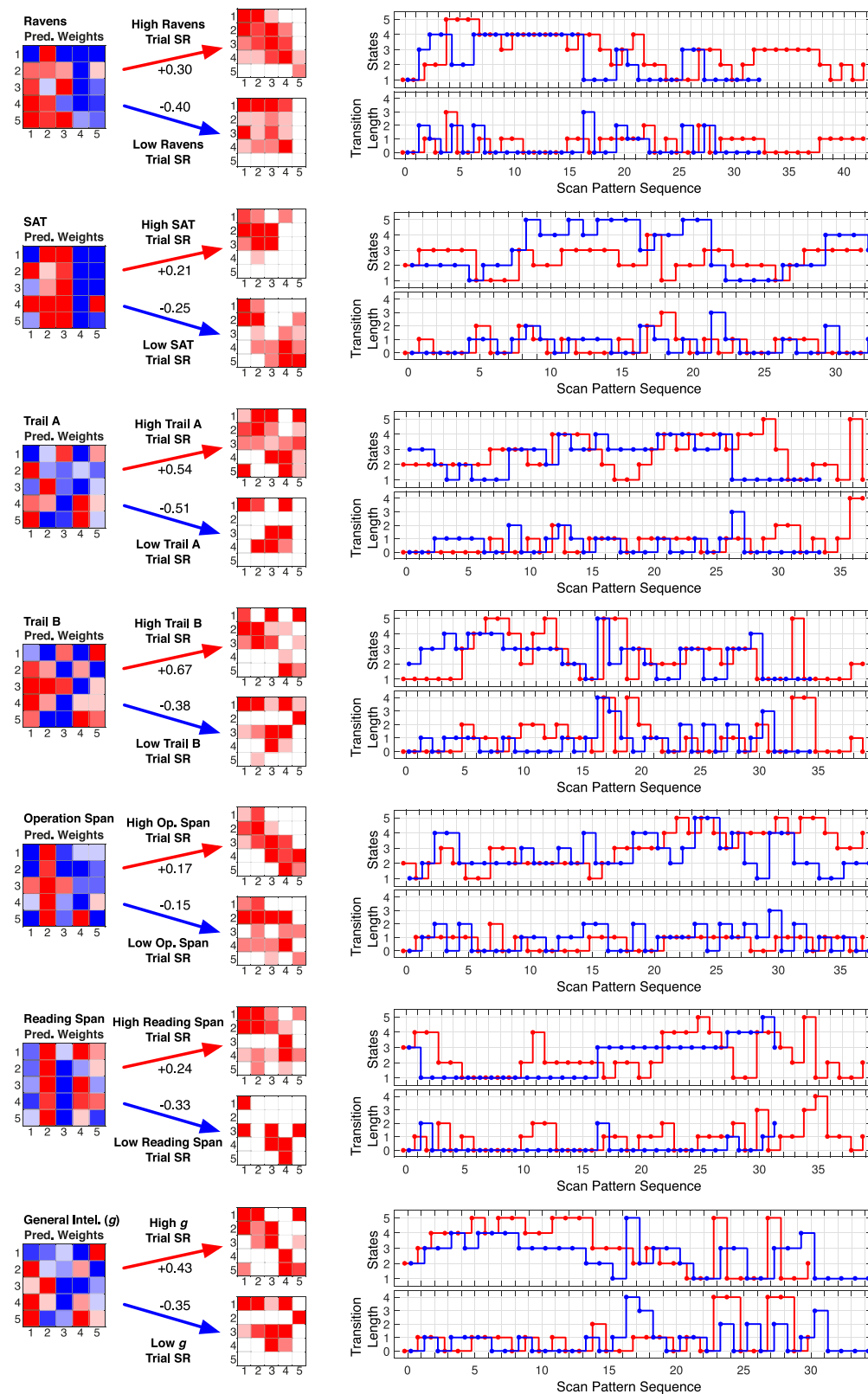


Figure 5. Illustrative positive and negative scan patterns for each individual difference SRSA model. For each individual difference measure the mean cross-validated prediction weights for the best SRSA model are shown. Illustrative trial scan patterns that were strongly positively/negatively correlated with the prediction weights and their corresponding trial SRs are shown to the right of each set of prediction weights. The top scan pattern panel shows the state transitions at each scan pattern position and the bottom scan pattern figure shows the transition length of each state transition. In the prediction weights and trial SR matrices the x-axis represents the sender state and the y-axis represents the receiver state.

Cognitive ID measures	Radiating state space				Vertical state space				Horizontal state space			
	$R^2$	$R^2_{cv}$	$\gamma$	$\alpha$	$R^2$	$R^2_{cv}$	$\gamma$	$\alpha$	$R^2$	$R^2_{cv}$	$\gamma$	$\alpha$
Raven's	0.66	0.43*	0.93	0.36	0.56	0.15	0.69	0.88	0.63	0.30	0.02	1.00
SAT	0.75	0.45*	0.61	0.16	0.78	0.50	0.35	0.39	0.73	0.20	0.01	0.01
Trail A	0.64	0.37	0.03	0.92	0.65	0.42*	0.49	0.51	0.64	0.29	0.73	0.71
Trail B	0.49	0.07	0.79	0.98	0.65	0.34*	0.55	0.68	0.55	0.18	0.86	1.00
Operation span	0.66	0.45*	0.50	0.14	0.65	0.32	0.04	0.93	0.60	0.31	0.23	0.14
Reading span	0.65	0.22	0.96	0.27	0.64	0.41*	0.10	0.20	0.61	0.28	0.79	0.62
General intelligence	0.56	0.09	0.75	0.98	0.69	0.41*	0.45	0.50	0.64	0.20	0.83	0.99

Table 1. Successor Representation Scanpath Analysis (SRSA) results: Goodness-of-fit  $R^2$  and leave-one-out cross-validation ( $R^2_{cv}$ ) for predicting individual differences (ID) in cognitive capacities from scan pattern regularities for all three state spaces (radiating, vertical, and horizontal). *Notes:* An asterisk highlights the SRSA models for each cognitive individual difference measure that are discussed in detail in the results section and shown in Figures 4 and 5.

state space explained 43% of the variance in individual differences in Raven's score (fluid intelligence) with a high SRSA gamma parameter ( $\gamma = 0.93$ ). Recall a higher gamma value means the relevant scan patterns are temporally elongated over many state time steps. The Raven's prediction weight matrix in Figure 4 indicated benefits to scene scan patterns that systematically moved between central and peripheral scene regions, with more expected visits to central regions (1 2 3) and fewer expected visits to peripheral regions (4 5). These scan pattern regularities can also be seen in the ideal Raven's example scan patterns shown in Figure 5. Previous work has shown that higher Raven's scores are associated with systematically processing Raven problem information (Hayes et al., 2011, 2015). These findings suggest that similar systematic information processing strategies may be employed by high fluid intelligence individuals when encoding complex real-world scenes. The radiating state space also explained 45% of the variance in individual differences in SAT score (crystallized intelligence) with a moderately high gamma parameter ( $\gamma = 0.61$ ). First, it is worth noting that the SAT prediction weights are very similar to the Raven's prediction weight matrices. Like the Raven's prediction weights, the SAT prediction weights show that higher SAT scoring individuals tended to spend more time systematically moving between central scene information, while lower scoring individuals more frequently visited peripheral scene information. The parity between the Raven's and SAT prediction weights suggests that general intelligence, containing constituent fluid and crystallized intelligences, may be associated with systematic information processing strategies for any complex visual display.

## Speed of processing

SRSA using the vertical state space explained more than a third of the variance in both speed of processing measures (Trail A and Trail B test scores) with similar

gamma values (Trail A,  $\gamma = 0.49$ ; Trail B,  $\gamma = 0.55$ ). As can be seen in Figure 4 the Trail A and Trail B prediction weight matrices resemble each other. Given the similarity between the Trail A and Trail B individual difference tasks and the common vertical state space, this is not surprising. The Trail A prediction weight matrix indicated that longer completion times were related to repeated same-state visits between the top (State 1) and bottom half of the display (States 3 and 4). This same pattern was present in the Trail B prediction weight matrix, but with fewer repeat fixations within a common vertical state, and instead more frequent/longer length state transitions across the full vertical extent of the scene. These pattern differences in the degree of repeating visits and transition length, may be related to the difference between the demands of the Trail A and Trail B tests. Specifically, the Trail A test only requires participants to sequentially link a single feature dimension (i.e., number), whereas the Trail B task requires keeping track and comparing two different dimensions (number and letter) across the visual field. However, more targeted work is needed to test this hypothesis generated by our exploratory SRSA analysis.

## Working memory capacity

There was no common state space definition that captured both operation span and reading span well. Individual differences in operation span were best predicted by the radiating SRSA model ( $R^2_{cv} = 0.45$ ) and reading span was best predicted by the vertical SRSA model ( $R^2_{cv} = 0.41$ ). The SRSA operation span model had a larger gamma parameter value ( $\gamma = 0.50$ ), while the SRSA reading span model had a small gamma ( $\gamma = 0.10$ ), which approximated a first-order transition matrix. The operation span prediction weight matrix indicated benefits to frequent visits to the region just outside of the display center (State 2) early in the scan pattern, followed by visits between the early

periphery (State 4) and the peripheral (State 5) scene information. The reading span prediction weight matrix with its small gamma value can effectively be interpreted as a first-order transition matrix and showed benefits to frequent transitions to the upper center region (State 2) and lower interior region (State 4). It is unclear why these two working memory capacity measures were not captured by a common state space. However, the act of reading sentences does involve clear attentional regularities, such as reading left to right and top to bottom. These regularities may be providing a boost to the vertical state space for reading span.

## General intelligence

Individual differences in general intelligence ( $g$ ) were best predicted by the vertical SRSA model ( $R^2_{cv} = 0.41$ ). The prediction weights are similar to the speed of processing models where longer trail completion times were related to greater numbers of expected visits between the top (State 1) and bottom half of the display (States 3 and 4). This similarity between the general intelligence prediction weights and the speed of processing prediction weights reflects the relatively high loading of the speed of processing measures (Trail A loading = 0.70; Trail B loading = 0.92) in the factor analysis used to estimate general intelligence (see Appendix B).

## Discussion

In this article, we used a powerful technique for scan pattern analysis to quantify the association between individual differences in scan patterns during scene viewing and individual differences in viewer cognitive capacity. Participants completed a scene encoding task while their eye movements were recorded, and a separate individual difference test battery that included measures of intelligence, working memory capacity, and speed of processing. SRSA (Hayes et al., 2011) was used to extract individual differences in participants' scene scan patterns that predicted individual differences in participants' cognitive capacities. The results revealed individual differences in scan patterns during scene encoding that explained more than 40% of the variance in viewer intelligence and working memory capacity measures, and more than a third of the variance in the speed of processing measures.

Most of the scene perception studies from the last 80 years that have reported eye movement scan patterns have noted that even when the scene stimulus and task are held constant, different participants often produce

qualitatively different scan patterns (Buswell, 1935; DeAngelus & Pelz, 2009; Henderson & Hollingworth, 1998; Noton & Stark, 1971a, 1971b; Underwood et al., 2009; Yarus, 1967). These earlier findings motivated us to investigate the relationship between individual differences in scan patterns and the cognitive individual differences of viewers. While previous studies have shown individual differences in fixation duration and saccade amplitude distributions during scene viewing (Andrews & Coppola, 1999; Castelhana & Henderson, 2008; Henderson & Luke, 2014), our findings are the first to provide evidence of a strong association between individual differences in scan patterns during scene viewing and cognitive individual differences among viewers.

Our results also have important theoretical implications for computational models of gaze control during complex visual tasks like scene viewing. As was discussed in the introduction, the majority of research and modeling of gaze control and overt attention has focused on how bottom-up stimulus features and top-down task goals influence gaze control during scene viewing, with relatively little work devoted to understanding the relationship between individual differences among viewers and gaze control (Castelhana & Henderson, 2008). Our findings arguably provide the strongest evidence to date that the underlying cognitive capacities of the viewer are also important for understanding gaze control during real-world scene viewing. Specifically, our results suggest that different dimensions of cognitive capacity are associated with different global information processing strategies during the encoding of scene information. These findings provide important new viewer-capacity constraints on models of gaze control during scene viewing, in addition to the image- and task-based cognitive factors that are typically considered in modeling gaze behavior.

Our results also revealed that scan patterns may be more informative than the location and/or duration of eye movements for understanding the role of viewer individual differences during scene viewing. For instance, a traditional eye metric model (see Appendix C) that ignored the sequential eye movement pattern information and only measured the overall duration, frequency, and distance between fixations was incapable of explaining any variance in the major cognitive capacity measures we collected. It was only when considering the sequential patterns between fixation locations that we were able to extract information on how gaze control was correlated with the underlying individual differences in the cognitive capacities of the viewers. Therefore, our results suggest not only that viewer properties are an important component of gaze control during scene encoding, they also suggest sequential scan patterns as a critical target measure to test future models of gaze control.



Finally, our results suggest that scan patterns during scene encoding could be applied to extract large amounts of individual difference data from a single, quick task. Our scene encoding task presented 40 scenes for 12 s each (8 min total viewing time), and from this limited data we were able to gain estimates of participants' fluid intelligence, crystallized intelligence, speed of processing, and working memory capacities. Specifically, SRSA was able to identify scan pattern regularities that accounted for between one third and one half of the variance in the cognitive capacities we measured. It is likely that larger sample sizes would improve the SRSA generalization performance even more by providing more finely tuned SR model parameter values.

While our study represents an initial step toward understanding the relationship between individual differences in scan patterns and individual differences in cognitive capacity during scene perception, it is also limited in a number of ways. First, while our data show a strong association between individual differences in scan patterns and cognitive individual differences, it remains an open question what is driving this association. One possibility is that the cognitive systems underlying individual differences in viewer's intelligence, working memory capacity, and speed of processing influence the functioning of the gaze control system. The opposite possibility is that individual differences in the gaze control system influence the functioning of the cognitive systems underlying intelligence, working memory capacity, and speed of processing. A third possibility is that the association is caused by an unknown third variable, such as a shared strategy between the scene encoding task and the cognitive tests. Of course, these possibilities are not mutually exclusive and the association we found could result from some mixture of these different explanations. Moreover, the underlying source of the association may be different for intelligence, working memory capacity, and speed of processing.

A second limitation of our study is that participants only completed a scene encoding task, so it is unclear whether task demands modulate the relationship between individual differences in scan patterns and the cognitive capacities we collected. It could be that the cognitive capacity measures we collected are specific to scene encoding, and may not generalize to other tasks (e.g., visual search) or to situations when there is no explicit task (e.g., free viewing). Third, we only examined cognitive measures of individual difference (i.e., intelligence, speed of processing, working memory capacity). There may be a number of other dimensions of individual differences (e.g., clinical measures or age) that are also relevant for scan patterns during scene viewing. For instance, there has been some work suggesting autism spectrum disorder and attention

deficit hyperactivity disorder can impact attentional control characteristics (Burgess et al., 2010; Remington, Swettenham, Campbell, & Coleman, 2009). Finally, we examined only three different potential state spaces (i.e., Radiating, Vertical, Horizontal) that defined and captured individual differences in broad scanning tendencies across all scenes. There may be more informative state space definitions.

The limitations of our current study suggest a number of promising avenues for future research. First, it would be informative to have participants complete multiple scene viewing tasks (e.g., encoding, visual search, and free viewing) to quantify how task goals interact with individual differences in scan patterns and viewer capacities. It may also be useful to expand the battery of individual difference measures to include clinical measures of individual difference that are known to be relevant for attentional control such as an index of attention deficit hyperactivity disorder or autism spectrum disorder. A study of these clinical measures could provide interesting new insights into atypical attentional control during scene viewing. Finally, in future work it would be useful to explore other conceptualizations of scanning behavior such as dynamic scene-specific state spaces that use either bottom-up saliency or observer-driven saliency (e.g., the five most looked at regions in each scene) to define the states within each scene. This could provide important information about how bottom-up image-based visual saliency interacts with individual differences in scan patterns and viewer capacities.

In summary, we found that individual differences in scan patterns during scene encoding predicted individual differences in participants' intelligence, speed of processing, and working memory capacity scores. Broadly, these findings suggest an important new link between individual differences in gaze control and individual differences in cognitive capacity when encoding real-world scene information. These quantitative results bridge the knowledge gap between earlier qualitative observations of individual differences in viewer scan patterns by demonstrating a strong association with individual differences in viewer cognitive capacities. Finally, our results offer important new viewer-capacity constraints on models of gaze control and suggest sequential scan patterns as a promising target measure. An important direction for future work will be trying to determine why gaze control and cognitive capacity are strongly associated, and integrating how individual differences in viewer capacities interact with bottom-up image-based properties and top-down task demands.

*Keywords:* scene viewing, eye movements, individual differences, intelligence, working memory capacity, speed of processing

## Acknowledgments

This research was supported by the National Science Foundation (BCS-1636586).

Commercial relationships: none.

Corresponding author: Taylor R. Hayes.

Email: trhayes@ucdavis.edu.

Address: Center for Mind and Brain, University of California, Davis, CA, USA.

## References

- Aigner, W., Miksch, S., Schumann, H., & Tominski, C. (2011). *Visualization of time-oriented data*. London: Springer.
- Andrews, T. J., & Coppola, D. M. (1999). Idiosyncratic characteristics of saccadic eye movements when viewing different visual environments. *Vision Research*, 39, 2947–2953.
- Borji, A., & Itti, L. (2014). Defending Yarbus: Eye movements reveal observers' task. *Journal of Vision*, 14(3):29, 1–22, doi:10.1167/14.3.29. [PubMed] [Article]
- Bors, D. A., & Stokes, T. L. (1998). Raven's Advanced Progressive Matrices: Norms for first-year university students and the development of a short form. *Educational and Psychological Measurement*, 58, 382–398.
- Bruce, N. D., & Tsotsos, J. K. (2009). Saliency, attention, and visual search: An information theoretic approach. *Journal of Vision*, 9(3):5, 1–24, doi:10.1167/9.3.5. [PubMed] [Article]
- Burgess, G. C., Depue, B. E., Ruzic, L., Willcutt, E. G., Du, Y. P., & Banich, M. T. (2010). Attentional control activation relates to working memory in attention-deficit/hyperactivity disorder. *Biological Psychiatry*, 67(7), 632–640.
- Buswell, G. T. (1935). *How people look at pictures*. Chicago: University of Chicago Press.
- Caffarra, P., Vezzadini, G., Zonato, F., Copelli, S., & Venneri, A. (2003). A normative study of a shorter version of Raven's Progressive Matrices 1938. *Neurological Sciences*, 24(5), 336–339.
- Castelhano, M. S., & Henderson, J. M. (2008). Stable individual differences across images in human saccadic eye movements. *Canadian Journal of Experimental Psychology*, 62(1), 1–14.
- Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, 54, 1–22.
- Churchland, P. S., Ramachandran, V. S., & Sejnowski, T. J. (1994). A critique of pure vision. In C. Koch & J. L. Davis (Eds.), *Large scale neuronal theories of the brain* (pp. 24–60). Cambridge, MA: MIT Press.
- Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychological Bulletin*, 12, 769–786.
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19(4), 450–466.
- Dayan, P. (1993). Improving generalization for temporal difference learning: The successor representation. *Neural Computation*, 5(4), 613–624.
- DeAngelus, M., & Pelz, J. B. (2009). Top-down control of eye movements: Yarbus revisited. *Visual Cognition*, 17(6), 790–811.
- Everitt, B. S., & Dunn, G. (2001). *Applied multivariate analysis*. New York: Oxford University Press.
- Findlay, J. M., & Gilchrist, I. D. (2003). *Active vision: The psychology of looking and seeing*. Oxford: Oxford University Press.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. New York: Springer.
- Hayes, T. R., Petrov, A. A., & Sederberg, P. B. (2011). A novel method for analyzing sequential eye movements reveals strategic influence on Raven's Advanced Progressive Matrices. *Journal of Vision*, 11(10):10, 1–11, doi:10.1167/11.10.10. [PubMed] [Article]
- Hayes, T. R., Petrov, A. A., & Sederberg, P. B. (2015). Do we really become smarter when our fluid-intelligence scores improve? *Intelligence*, 48(1), 1–14.
- Haykin, S. (2009). *Neural networks and learning machines* (3rd ed.). New York: Prentice Hall.
- Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, 7(11), 498–504.
- Henderson, J. M., & Hollingworth, A. (1998). Eye movements during scene viewing: An overview. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 269–293). Oxford: Elsevier.
- Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology*, 50, 243–271.
- Henderson, J. M., & Luke, S. G. (2014). Stable individual differences in saccadic eye movements during reading, pseudo-reading, scene viewing, and

- scene search. *Journal of Experimental Psychology: Human Perception and Performance*, 40(4), 1390–1400.
- Henderson, J. M., Malcolm, G. L., & Schandl, C. (2009). Searching in the dark: Cognitive relevance drives attention in real-world scenes. *Psychonomic Bulletin & Review*, 16, 850–856.
- Henderson, J. M., Shinkareva, S. V., Wang, J., Luke, S. G., & Olejarczyk, J. (2013). Predicting cognitive states from eye movements. *PloS One*, 8(5), 1–6.
- Holmqvist, K., Nyström, M., & Mulvey, F. (2012). Eye tracker data quality: What it is and how to measure it. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (pp. 45–52). New York: ACM.
- Holmqvist, K., Nyström, R., Andersson, M., Dewhurst, R., Jorodzka, H., & van de Weijer, J. (2015). *Eye tracking: A comprehensive guide to methods and measures*. Oxford: Oxford University Press.
- Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, 40, 1489–1506.
- Itti, L., & Koch, C. (2001). Computational modeling of visual attention. *Nature Reviews Neuroscience*, 2, 194–203.
- Jensen, A. R. (1998). *The g factor: The science of mental ability*. London: Praeger.
- Kardan, O., Berman, M. G., Yourganov, G., Schmidt, J., & Henderson, J. M. (2015). Classifying mental states from eye movements during scene viewing. *Journal of Experimental Psychology: Human Perception and Performance*, 41, 1502–1514.
- Koch, C., & Ullman, U. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology*, 4, 219–227.
- Koehler, K., Guo, F., Zhang, S., & Eckstein, M. P. (2014). What do saliency models predict? *Journal of Vision*, 14(3):14, 1–27, doi:10.1167/14.3.14. [PubMed] [Article]
- Manual of Directions and Scoring*. (1944). Washington, DC: War Department, Adjutant General's Office.
- Martin, E. (1974). Saccadic suppression: A review and an analysis. *Psychological Bulletin*, 81(12), 899–917.
- Noton, D., & Stark, L. (1971a). Scanpaths in eye movements during pattern perception. *Science*, 171, 308–311.
- Noton, D., & Stark, L. (1971b). Scanpaths in saccadic eye movements while viewing and recognizing patterns. *Vision Research*, 11, 929–942.
- Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research*, 42, 102–123.
- Picard, R. R., & Cook, D. R. (1984). Cross-validation of regression models. *Journal of the American Statistical Association*, 79(387), 575–583.
- Raven, J. C., Raven, J., & Court, J. H. (1998). *Manual for Raven's progressive matrices and vocabulary scales. Section 4: Advanced progressive matrices*. San Antonio, TX: Pearson.
- Reitan, R. M., & Wolfson, D. (1985). *The Halstead-Reitan Neuropsychological Test Battery: Therapy and clinical interpretation*. Tucson, AZ: Neuropsychological Press.
- Remington, A., Swettenham, J., Campbell, R., & Coleman, M. (2009). Selective attention and perceptual load in autism spectrum disorder. *Psychological Science*, 20(11), 1388–1393.
- Salthouse, T. A. (2011). What cognitive abilities are involved in trail-making performance? *Intelligence*, 39(4), 222–232.
- SAT. (2014). New York: The College Board.
- SR Research. (2010a). *Experiment Builder user's manual*. Mississauga, ON: Author.
- SR Research. (2010b). *EyeLink 1000 user's manual, version 1.5.2*. Mississauga, ON: Author.
- Stark, L., & Ellis, S. R. (1981). Scanpath revisited: Cognitive models of direct active looking. In D. F. Fisher, R. A. Monty, & J. W. Senders (Eds.), *Eye movements: Cognition and visual perception* (pp. 193–226). Hillsdale, NJ: Lawrence Erlbaum.
- Sutton, R. S. (1988). Learning to predict by the methods of temporal differences. *Machine Learning*, 3(1), 9–44.
- Tatler, B. W. (2007). The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision*, 7(14):4, 1–17, doi:10.1167/7.14.4. [PubMed] [Article]
- Thiele, A., Henning, P., Kubischik, M., & Hoffman, K. P. (2002). Neural mechanisms of saccadic suppression. *Science*, 295(5564), 2460–2462.
- Torrallba, A. (2003). Modeling global scene factors in attention. *Journal of the Optical Society of America, A: Optics, Image Science, & Vision*, 20, 1407–1418.
- Torrallba, A., Oliva, A., Castelano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review*, 113, 766–786.
- Underwood, G., Foulsham, T., & Humphrey, K. (2009). Saliency and scan patterns in the inspection of real-world scenes: Eye movements during encoding and recognition. *Visual Cognition*, 17, 812–834.



- Vigneau, F., Caissie, A. F., & Bors, D. A. (2006). Eye-movement analysis demonstrates strategic influences on intelligence. *Intelligence*, 34(3), 261–272.
- Yarbus, A. L. (1967). *Eye movements and vision*. New York: Springer.

## Appendix A

### Determination of sample size

Target sample size was determined on the basis of previous work using scan patterns and Successor Representation Scanpath Analysis to predict individual differences in cognitive capacity (Hayes et al., 2011, 2015), which indicated that a relatively small number of participants ( $N = 35$ ) can be used to train and validate a predictive Successor Representation Scanpath Analysis model. Therefore, we aimed to collect eye movement data and cognitive capacity data so that each cognitive capacity measure had approximately 35 participants. Since all of our analyses were focused on quantifying how well scan patterns (and other eye metrics) can predict cognitive capacities, the main risk that must be accounted for in the predictive models based on these samples is overfitting. In order to control for overfitting and provide estimates of how well each predictive model will generalize to new data, leave-one-out cross-validation (Haykin, 2009; Picard & Cook, 1984) was performed and is reported for all analyses.

## Appendix B

### Individual difference measures

Intelligence was measured using two different tests: a short form of the Advanced Raven's Progressive Matrices Test (RAPM; Raven, Raven, & Court, 1998) and the SAT test (SAT, 2014). RAPM is a visual geometric analogy test that measures novel problem solving ability, known as *fluid intelligence* (Cattell, 1963; Jensen, 1998). A Raven problem consists of a matrix and eight response alternatives. There are multiple distinct relations among the entries in a given row or column of the matrix, and the participants had to identify the relations and select the response that best matched the pattern. A short form of the RAPM (Bors & Stokes, 1998; Caffarra, Vezzadini, Zonato, Copelli, & Venneri, 2003) was administered and fluid intelligence was measured as the total number of Raven items answered correctly. Participants' self-reported SAT scores were used as an index of their knowledge

gained through experience, or *crystallized intelligence* (Cattell, 1963; Jensen, 1998). Therefore, together the intelligence battery contained individual difference measures of both major components of general intelligence (i.e., fluid and crystallized intelligence).

Working memory capacity was measured using two different complex span tasks: an operation span task and a reading span task. The complex operation span task required participants to remember a sequence of letters that are each followed with an arithmetic processing task (Conway et al., 2005; Salthouse, 2011). Capacity in the operation span task is measured as a function of the number of letters that can be correctly recalled. In the reading span task participants read a series of sentences and had to recall the last word of each sentence in order of appearance (Daneman & Carpenter, 1980). Reading span was measured as the number of last words correctly recalled in order. The operation and reading complex span tasks have both been widely used to measure individual differences in working memory capacity (for review see Conway et al., 2005).

Speed of processing was measured using the Trail Making Test A and B (*Manual of Directions and Scoring*, 1944; Reitan & Wolfson, 1985). The Trail Making Test A and B involved timing how long it took participants to draw a line between a jumbled set of sequentially numbered dots (Trail A) and a sequence of sequential numbers and letters (Trail B). The Trail Making Test is thought to measure a number of factors including speed of processing and executive control (Salthouse, 2011). It is worth noting that the Trail A and B measures are the only included cognitive capacity measures where higher scores indicated lower cognitive ability, since each task is measured as a function of completion time, where faster times indicate greater capacity. For the sake of consistency of interpretation across all six cognitive capacity measures (i.e., higher scores indicate greater capacity), the Trail A and Trail B scores were renormalized by multiplying the standardized data by  $-1$  so that higher scores indicated faster completion times.

Finally, general intelligence ( $g$ ) was estimated for each participant using factor analysis. Excluding SAT scores, there were 51 participants that completed the remaining five measures (Raven's, Trail A, Trail B, operation span, and reading span measures). A factor analysis was performed, and the first factor from the unrotated solution was used as an index of  $g$ . The estimated loadings revealed the largest weighting for speed of processing measures (Trail A = 0.70; Trail B = 0.92), followed by fluid intelligence (Raven's = 0.46) and working memory capacity (Operation span = 0.22; reading span = 0.63).

Cognitive ID measures	$R^2$	$R^2_{cv}$	RMSE	RMSE <sub>cv</sub>
Raven's	0.09	0.01	2.64	2.95
SAT	0.11	0.01	291.51	334.19
Trail A	0.06	0.03	5.57	6.15
Trail B	0.11	0.01	12.45	13.95
Operation span	0.09	0.01	16.58	19.06
Reading span	0.07	0.06*	0.82	0.96
General intelligence	0.14	0.01	0.87	0.99

Table C1. Goodness-of-fit and leave-one-out cross-validated performance for predicting individual difference measures using traditional eye metrics. *Notes:* The traditional eye metric model included mean fixation duration, saccade amplitude, and fixation number as predictors in a multiple regression model to predict each individual difference measure. The results revealed that traditional eye metrics are not able to predict any of the underlying individual differences we measured. \*Note this correlation value is driven by a spurious outlier, the cross-validated model prediction is actually worse as reflected in the increase in the root mean squared error (RMSE).

## Appendix C

### Traditional and transition probability models

SRSA performance was compared to two simpler eye movement data models: a traditional eye metric model and a first-order transition probability model. The traditional eye metric model calculated the mean fixation duration, saccade amplitude, and fixation number for each participant across all 40 scenes. These three eye metrics were then used as predictors in a multiple regression model to predict each individual difference measure. To allow a direct comparison between the traditional eye metric model performance and the SRSA model performance, both a goodness-of-fit  $R^2$  and leave-one-out cross-validation ( $R^2_{cv}$ ) were computed for each cognitive capacity measure.

The transition probability model computed first-order transition probabilities for the same three state

spaces that were used for each SRSA model (i.e., radiating, vertical, and horizontal). The only difference between the transition probability models and the SRSA models is that a first-order transition matrix was computed for each trial scan pattern rather than a successor representation (SR). The same temporal difference learning rate parameter  $\alpha$ , dimensionality reduction using PCA, and cross-validation procedures were performed for the first-order transition probability models. The comparison between the SRSA models and the transition probability models provides a direct estimate of the gains in prediction performance that are due to the temporal difference learning rate and PCA dimensionality reduction, versus the performance gains that are due to the power of the SR to extract temporally extended scan pattern regularities beyond first-order transitions.

The traditional eye movement model results for each individual difference measure are shown in Table C1. The traditional eye metric model used mean fixation duration, saccade amplitude, and fixation number as predictors and was not able to account for a significant amount of variance in any of the cognitive individual difference measures we collected. Table C2 shows the performance of the first-order transition probability model that used an identical prediction algorithm to the SRSA models. The only difference between the SRSA model and transition probability model was instead of using an SR for each scan pattern, a first-order transition probability was computed instead. While the first-order transition probability model was able to successfully explain some of the variance in individual difference measures, a comparison with SRSA revealed that on average SR increased generalization performance by 67% (median = 55%). These results highlight the large benefit of extending the temporal horizon beyond just first-order transitions by using SR to capture temporally extended regularities in scan patterns. Finally, it is worth noting the importance of performing cross-validation to test models of eye movement data that is

Cognitive ID measures	Radiating state space			Vertical state space			Horizontal state space		
	$R^2$	$R^2_{cv}$	$\alpha$	$R^2$	$R^2_{cv}$	$\alpha$	$R^2$	$R^2_{cv}$	$\alpha$
Raven's	0.50	0.04	0.15	0.53	0.06	0.03	0.63	0.26	1.00
SAT	0.61	0.33	0.81	0.66	0.27	0.17	0.72	0.20	0.01
Trail A	0.64	0.29	0.93	0.59	0.23	0.57	0.46	0.02	0.39
Trail B	0.46	0.03	0.28	0.63	0.27	0.07	0.55	0.07	0.99
Operation span	0.62	0.10	0.50	0.65	0.27	0.92	0.54	0.17	0.72
Reading span	0.59	0.20	0.07	0.64	0.40	0.20	0.56	0.04	0.65
General intelligence	0.49	0.04	0.24	0.69	0.38	0.09	0.65	0.15	0.99

Table C2. First order transition probability results: Goodness-of-fit  $R^2$  and leave-one-out cross-validated ( $R^2_{cv}$ ) for predicting individual differences (ID) in cognitive capacities from scan patterns using first-order transition probability instead of successor representation. *Notes:* A comparison with the SRSA performance in Table 1 shows that successor representation provides an average increase in generalization performance ( $R^2_{cv}$ ) of 67% (median 55%) relative to first-order transition probabilities.

evident in Tables 1, C1, and C2. Consistent with previous modeling of eye movement data (Hayes et al., 2011, 2015; Vigneau, Caissie, & Bors, 2006), the SRSA, transition probability, and tradition eye metric model results all show the goodness-of-fit  $R^2$  is

consistently inflated due to overfitting. Our results support the general recommendation that statistical models of eye movement data should be cross-validated to provide more accurate estimates of their ability to generalize to new data.