Exploring Individual Differences via Clustering Capacity Coefficient Functions

Joseph W. Houpt (joseph.houpt@wright.edu)
Wright State University, Department of Psychology, 3649 Colonel Glenn Highway, Dayton, OH 45435 USA

Leslie M. Blaha (leslie.blaha@us.af.mil)
711[th] Human Performance Wing, Air Force Research Laboratory, 2255 H Street, Wright-Patterson AFB, OH 45433 USA

Abstract

The capacity coefficient function is a well-established, model-based measure comparing performance with multiple sources of information together to performance on each of those information sources in isolation. Because it is a function across time, it may contain a large amount of information about a participant. In many applications, this information has been ignored, either by using qualitative assessment of the function or by using a single summary statistic. Recent work has demonstrated the efficacy of functional principal components analysis for extracting important information about the capacity function. We extend this work by applying clustering techniques to examine individual capacity differences in configural learning.

Keywords: Configural Learning; Individual Differences; Capacity Coefficient; Human Information Processing Modeling

Introduction

A critical facet of characterizing the cognitive mechanisms involved human information processing is capturing changes as information sources change. These models are applied to individual participant data, so they have strong potential to indicate individual differences. However, because of the functional nature of many information processing modeling approaches, it is challenging to find meaningful ways to aggregate the individual analyses to identify group trends while accounting for the variability of the processes of interest. We implement a rigorous approach to quantifying individual differences and group patterns in a functional statistic measuring the processing of multiple information sources, the capacity coefficient function (CCF). We then employ clustering techniques to capture both qualitative and quantitative trends in CCF data from a configural learning study (Blaha, 2010).

CCF analysis models the effects of changing information demands on an individual's information throughput. Information throughput refers to how much information can be processed in a given amount of time. The CCF is a model-based analysis comparing response times (RTs) in a task with multiple sources of information to RTs from task with a single isolated source of information (Townsend & Wenger, 2004; Houpt, Blaha, McIntire, Havig, & Townsend, 2014). There are a number of factors known to change performance as the number of information sources increases. These include correlated processing of the sources, facilitation/inhibition among the processes, processing strategy, and task demands. The CCF controls for the effects of task demands by utilizing a baseline prediction from an unlimited capacity, independent parallel (UCIP) model. UCIP predictions depend on the stopping rule for the task, and in this paper we will model exhaustive (AND) and a single-target self-terminating (ST-ST) tasks.

We use the cumulative reverse hazard function, defined by $K_i(t) = \ln[F_i(t)]$, where for channel i, $F_i(t) = P(T_i \leq t)$. $K(t)$ is interpreted as the amount of work left to be done by the system after t time has passed. In a system with more efficient throughput, the $K(t)$ increases faster to reach 0 earlier than a less efficient system.

In an UCIP model, individual sources of information are processed simultaneously without statistical interactions, and the addition of more sources never helps nor hurts the processing speeds of other sources. Under AND stopping, all sources must be completely processed before a response can be made. The exhaustive UCIP model predicts that the information throughput during the joint processing of multiple sources is the sum of their individual throughput measures. The CCF compares this prediction for a set of n information sources to the observed RT distribution from a condition requiring the simultaneous processing of all n sources together, as defined by Eq. 1.

$$C_{\text{and}}(t) = K_n(t) - \sum_{i=1}^{n} K_i(t) \tag{1}$$

AND processing is often juxtaposed with ST-ST processing. This is when target information is presented either alone or among distracting information. For ST-ST, the UCIP prediction is that the information throughput for the target source, i, will be the same regardless of whether or not there are additional, non-target sources of information presented. The CCF for ST-ST processing is

$$C_{\text{stst}}(t) = K_{\text{iX}}(t) - K_i(t). \tag{2}$$

CCFs are interpreted relative to 0. If $C(t) = 0$, then observed throughput is equal to the UCIP model and unlimited capacity is inferred. $C(t) > 0$ indicates super capacity processing, i.e., additional information sources resulted in more efficient performance. $C(t) < 0$ indicates limited capacity, wherein additional sources resulted in less efficient performance. Examples of the $C(t)$ classes are shown in Figure 1.

In early CCF applications, analysis was limited to visually assessing the function. Houpt and Townsend (2012) recently derived statistics for testing a null hypothesis of UCIP processing. While this statistic improves upon visual-only assessment, it marginalizes information about the shapes of the CCF.

Burns, Houpt, Townsend, and Endres (2013) demonstrated the use of functional principal components analysis (fPCA; Ramsay & Silverman, 2005) for analyzing differences in the

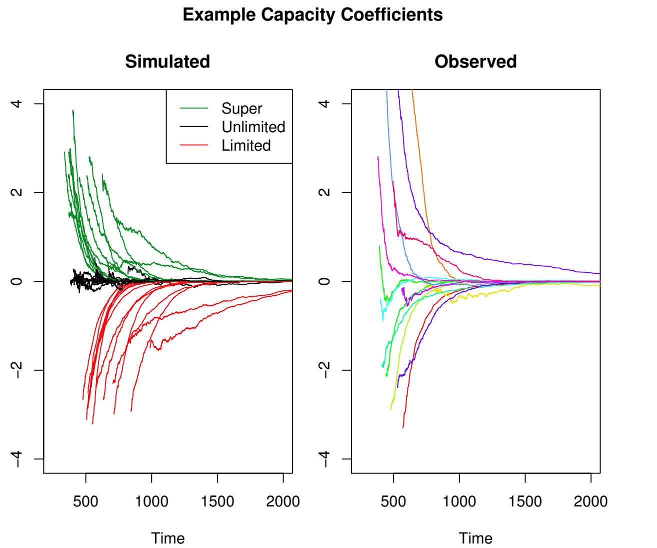

Figure 1: Example capacity coefficients. The left plot demonstrates data simulated from super, unlimited and limited capacity generating models. The right plot demonstrates capacity functions estimated from participant data (Day 4 of the configural learning study).

forms of the capacity functions. fPCA is similar to standard PCA. In PCA, the data are described as a linear combination of orthogonal vectors which are ordered by the amount of variance in the data along that vector. In fPCA, the data are described as a linear combination of orthogonal functions ($\int f_i f_j = 0$) which are ordered by the amount of variance in the data along that function (i.e., f_i maximizes $\sum_{k=1}^{N} \left(\int f_i x_k\right)^2$ where x_k are the observed functions, subject to the constraint that $\int f_i f_j = 0$ for $j < i$).

fPCA and PCA are often used to describe a dataset with a dimensional subspace than the original data by only using the first n bases (effectively projecting the data onto a lower dimensional subspace). Each individual datum is described by its factor scores on those n bases. For example, if $x_i = a_1 f_1 + a_2 f_2 \ldots a_m f_m$ where the f_j are the basis functions from fPCA, then we can use a lower dimensional representation of x_i given by $x_i \approx a_1 f_1 + a_2 f_2$. fPCA reduction can provide us with a tractable vector space together with representative functions to describe CCF data. In particular, similarity in the vector fPCA score space captures similarity in CCF shapes, thereby providing a way to quantify properties of the full functions. Thus, fPCA quantifies the shapes among a set of CCFs derived from individual participants, and we can further analyze the fPCA weights to identify trends among the individual differences.

Clustering

We applied two popular clustering methods to the fPCA-reduced capacity coefficients. Our goal is systematically and quantifiably capturing patterns of similarity and differences in CCFs only previously described in qualitative ways. K-means clustering refers to a technique in which a set of points (in any finite dimensional vector space) are modeled as members of one of K different clusters. The free parameters of the model are the locations of the center of each of the K clusters, chosen to minimize the Euclidean distance between each datum and its nearest cluster center. The number of clusters to use, K, is experimenter-specified, either using a scree plot or comparing the ratio of within cluster variation to between cluster variation across different values of K. We employ k-means clustering to determine the number of unique shapes among a set of CCFs.

Hierarchical clustering builds successively more inclusive groupings of data (agglomerative) or successively dividing the data into more exclusive groupings (divisive). We use a basic agglomerative procedure which first clusters the closest nodes. The next cluster is formed by either grouping a different pair of nodes which have the next smallest distance between them or by clustering a datum with the previously formed cluster if the distance between the datum and the cluster is less then the distance between any pair of data. This procedure iterates until a single cluster forms. We use hierarchical clustering to examine group trends emerging from individual participant CCF modeling.

Configural Learning Data

We analyzed the data from a configural learning study by Blaha (2010) in which the CCFs qualitatively changed over the course of training. Configural learning is the process by which individual object features are unified into a single perceptual unit. Configural learning through unitization changes the perceptual representation of the objects, and Blaha and colleagues demonstrated that this not only changes the information throughput supporting object classification (Blaha, 2010) but also changes the supporting scalp-level neural responses (Blaha, Busey, & Townsend, 2009).

The study entailed two categorization tasks based on Goldstone (2000). A conjunctive categorization task was designed to require AND processing of the category 1 object by systematic variation of the category 2 object features. Mandatory AND processing encouraged participants to chunk the features into a single object; thus, we expected (and previously observed) unitization of this object. Unitization increases information throughput over the course of learning, captured by CCFs shifting from limited to super capacity over training. A single-feature categorization task served as a baseline estimate for the UCIP predictions. Each category in this task only contained a single object, with one feature differing between the two objects. Thus, RTs in this task captured the speed of responding as participants learned to distinguish individual visual features. A total of fourteen participants completed 10-14 experimental sessions, including 5-7 training sessions of both the conjunctive and single-target categorization tasks. Each one-hour session consisted of 1200 trials. In all, the statistical learning herein utilized 12,000-

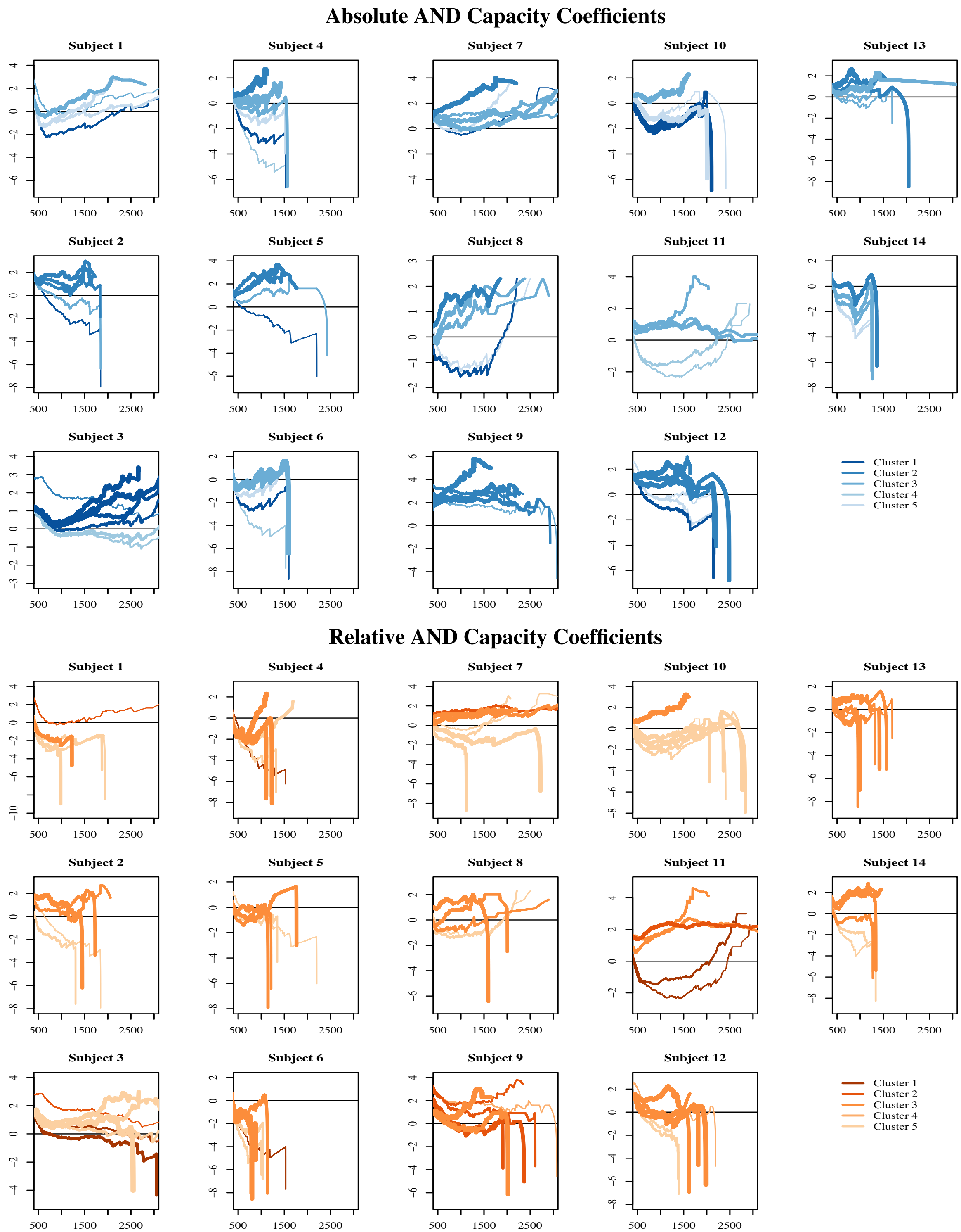

Figure 2: AND Capacity coefficients for all participants over all training. The upper half gives the absolute $C_{\text{and}}(t)$ data, and the lower half gives the relative $C_{\text{and}}(t)$ data. The thickness of each line indicates the training session where the thinnest lines are the first session and the thickest line in each plot is that participant's final day of training. Line colors indicate the $K = 5$ K-means cluster assignment for each $C_{\text{and}}(t)$ curve.

16,800 trials for each of the 14 participants (see Blaha, 2010, for full study details).

For every day of training, four CCFs were estimated for each participant. First, based on the mandatory AND stopping rule, the unitized object was examined with $C_{\text{and}}(t)$. The complementary non-conjunctive responses required the identification of features unique to category 2, engaging an ST-ST response rule. Thus, category 2 RTs were analyzed with $C_{\text{stst}}(t)$. For both $C_{\text{and}}(t)$ and $C_{\text{stst}}(t)$, absolute and relative capacity coefficients were estimated. Absolute learning measured changes in the CCFs with the UCIP estimate derived from the first training day, to give an overall estimate of capacity improvement from the start of the learning process. Relative learning varied the UCIP estimate for each day, to account for the single-target discrimination learning occurring in parallel with configural learning.

Figure 2 illustrates the AND CCF data for all participants. Day 1 of training is shown in the thinnest line, and the last day of training is the thickest in each plot. All participants exhibited $C_{\text{and}}(t)$ improvements over training, but as Figure 2 highlights, there was a variety of individual differences observed by Blaha (2010). For example, Subject 4 exhibited a gradual improvement from limited $C_{\text{and}}(t) < 0$ to super $C_{\text{and}}(t) > 0$, whereas Subjects 8 and 11 showed a more step-like shift from limited directly to super capacity. Subjects 9, 3, and 12 had strong speed-accuracy trade-offs, with super capacity early in training at the cost of lower accuracy. By applying unsupervised learning to systematically determine the numbers of unique patterns in the data we can quantify these verbal descriptions of the various learning patterns that would otherwise be merely observational inferences.

K-means Clustering

For all four CCF estimates, we extracted the first 4 fPCs, creating four 4-dimensional vector spaces in which we could compare the capacity data. fPCA analysis was performed separately for each of the four types of CCFs, and so we first analyze the unique components within each of the those $C(t)$ classes. K-means analysis was used to identify the number of unique $C(t)$ function shapes exhibited within each condition.

In all conditions, $K = 5$ clusters of functions fit the data. Figure 3 illustrates the five clusters for both the absolute (top) and relative (bottom) $C_{\text{and}}(t)$ fPCA score spaces, with similar results for $C_{\text{stst}}(t)$ fPCA scores. The CCFs plotted in Figure 3 illustrate the $C_{\text{and}}(t)$ functions representative of the centroids of each cluster. These were computed by $x_i \approx a_1 f_1 + a_2 f_2 + a_3 f_3 + a_4 f_4$ where $\{a_1, \ldots, a_4\}$ are the 4D centroid score values.

The shapes of the centroid CCFs (Figure 3) are consistent with the generally observed trends over learning. One cluster shows strict limited capacity $C_{\text{and}}(t) < 0$ values, consistent with the inefficient performance early in training. Other clusters show mixed values above and below 1, reflecting the functions in the middle of training that tend to shift from limited to unlimited to super capacity, as well as often showing non-flat shapes (e.g., super capacity for fast RTs, limited capacity for slow RTs). A final cluster exhibits strict super capacity $C_{\text{and}}(t) > 0$ values, consistent with participants reaching highly efficient throughput by the end of training. K-means clustering shows that the raw CCF data in configural learning can be classified into 5 fundamental shapes.

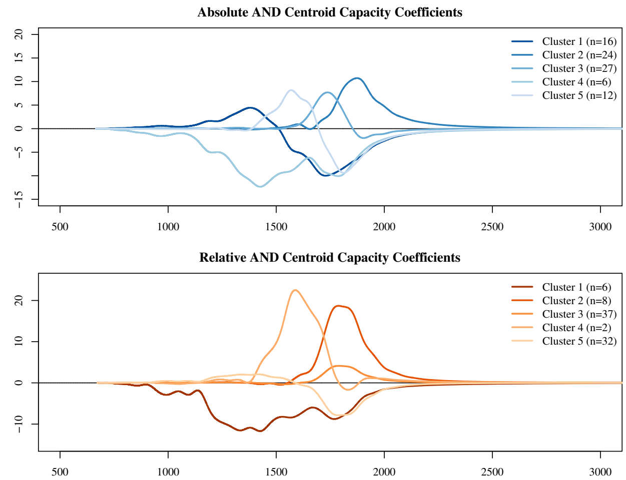

Figure 3: Representative $C_{\text{and}}(t)$ functions for each K-means cluster in both the absolute (upper) and relative (lower) fPCA-reduced capacity spaces. Centroid $C_{\text{and}}(t)$ functions determined by the linear combination of the centroid scores and the fPCs for each space.

Hierarchical Clustering

In order to look for groupings among the learning trends, we mapped the functional learning traces into a high-dimensional linear space by aggregating each participant's fPCA scores over all days of training. Participants exhibiting similar functional learning traces are represented by vectors close in this space. Note that because fPCA scores further represent a standardization of CCFs, we can map both the $C_{\text{and}}(t)$ and $C_{\text{stst}}(t)$ learning into the same high-dimensional space. Hierarchical clustering was performed on 20-dimensional fPCA score space. In this space, each participant was represented by four vectors, one for each type of CCF (relative and absolute $C_{\text{and}}(t)$ and $C_{\text{stst}}(t)$). The 20D vectors contained the four fPC weights over five days of training (the first five days if a participant trained longer). Distance between vectors, D, was estimated with the Euclidean metric. A heatmap of D is shown in Figure 4, with the rows ordered according to the hierarchical clustering results. Agglomerative clustering on D was performed with Ward's minimum variance method, minimizing total within-cluster variance (Ward, 1963).

Figure 5 depicts the dendrogram resulting from the hierarchical clustering analysis. It is immediately obvious that there is a clear division in fPCA score space between the data from the STST and AND conditions. The red bounding boxes illustrate a cut tree with four groupings. Note that increasing the number of groups in the cut tree further divided the $C_{\text{and}}(t)$ half of the dendrogram, leaving the $C_{\text{stst}}(t)$ portion clustered into a single group. One group contains all

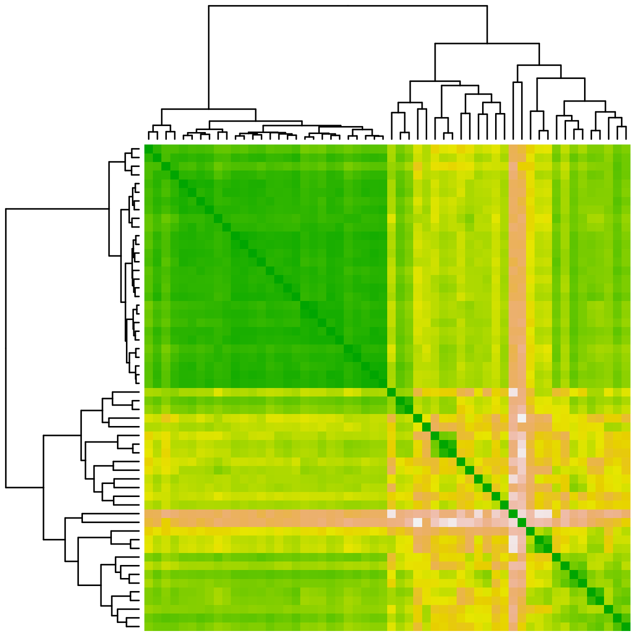

Figure 4: Heatmap of Euclidean distances in the 20D fPCA scores space. The columns are ordered according to the dendrogram depicted in the margins from the Wald hierarchical clustering. Green coloring gives smaller distances; white gives the largest distances.

the $C_{\text{stst}}(t)$ fPCA vectors, indicating that all participants exhibited similar changes in $C_{\text{stst}}(t)$ over the course of learning. The D values illustrated in Figure 4 confirm that all $C_{\text{stst}}(t)$ fPCA scores were highly similar.

The AND scores were split into three groups. Subject 9 separates early into her own cluster (confirmed by pairwise D values at the high end of the range), reflecting a unique pattern of AND learning different from all other participants. As illustrated in Figure 2, Subject 9 exhibited a unique combination of an increasing, strictly super absolute capacity learning curve with a U-shaped relative capacity learning pattern, resulting from a strong speed-accuracy trade-off.

The second AND cluster (dendrogram middle) contains the majority of the absolute $C_{\text{and}}(t)$ results. This indicates that in fPCA-reduced space, most participants exhibited similar learning-based changes in their CCFs. From Figure 2, we can see similarities in their profiles based on the K-means clustering of the individual curves. The colors in Figure 2 illustrate that participants in this cluster have CCFs landing in all 5 clusters. That is, their learning trajectories move through all the average function shapes illustrated in the absolute capacity centroid AND CCFs of Figure 3. This cluster also contains a subgroup of relative fPCA score vectors, which cluster with each other before clustering with the AND vectors. Similarly, the final AND cluster contained mostly relative capacity fPCA score vectors. Referring back to the K-means color coding for the relative capacity coefficients in Figure 2, this cluster represents those participants whose learning trajectories, measured by relative capacity, contain at least one function falling into all the CCF shapes illustrated by the relative centroid functions in the lower part of Figure 3. Again, there is a small subgroup of absolute capacity vectors that clustered into a similar part of 20D fPCA-reduced space.

We note that the small subgroup of relative (absolute) capacity scores clustering with the majority of the absolute (relative) capacity scores doesn't mean the original CCFs were the same between these two groups. This is because the fPCA scores were derived separately on the two types of capacity measures, and the weights in the fPCA-reduced space refer to different fPCs. What is important is that the separation of these small subgroups implies that whether measured in absolute or relative terms, configural learning can be supported by two different learning trends (three if you count the trend of Subject 9). With few exceptions in this data set, the subgroups consist of individual participants whose absolute and relative capacity measures clustered into similar portions of the fPCA-reduced space. Further analysis is needed to understand how this relates to similarity between the fPCs and other functional measures.

Discussion

In this paper we have demonstrated the use of clustering techniques to find group trends among individual differences in configural learning. The CCF gives a model-based measure of how people are using the information sources together without making specific assumptions about the RT distributions. Although the raw functions may be unwieldy for exploring sub-groups of participants, fPCA can be used to capture the important variation across CCFs. We then used standard clustering techniques to examine different performance patterns. The cluster memberships attained with these methods can either be used for additional exploratory analysis or for further comparisons with other types of data (e.g., clinical diagnosis or working memory capacity). Importantly, clustering and other statistical learning approaches can provide principled methods for finding generalizable patterns or trends in individual data without losing the characteristics in the individual participant data, which can be particularly challenging for functional or time series data.

In previous CCF applications, analysis had been confined to either qualitative, verbal descriptions of different patterns across capacity functions or to an aggregate statistic from Houpt and Townsend (2012). The approach presented in this paper allowed us to objectively identify different patterns across participants using the full functional information from the CCFs. From this we are able to conclude that configural learning requires at least five unique CCF shapes to describe all the observed stages of learning captured in $C(t)$ functions. Each participant fell into one of three learning patterns identified by hierarchical clustering. So while all participants unitized the objects in the task and showed overall information throughput increases, there were three different trajectories

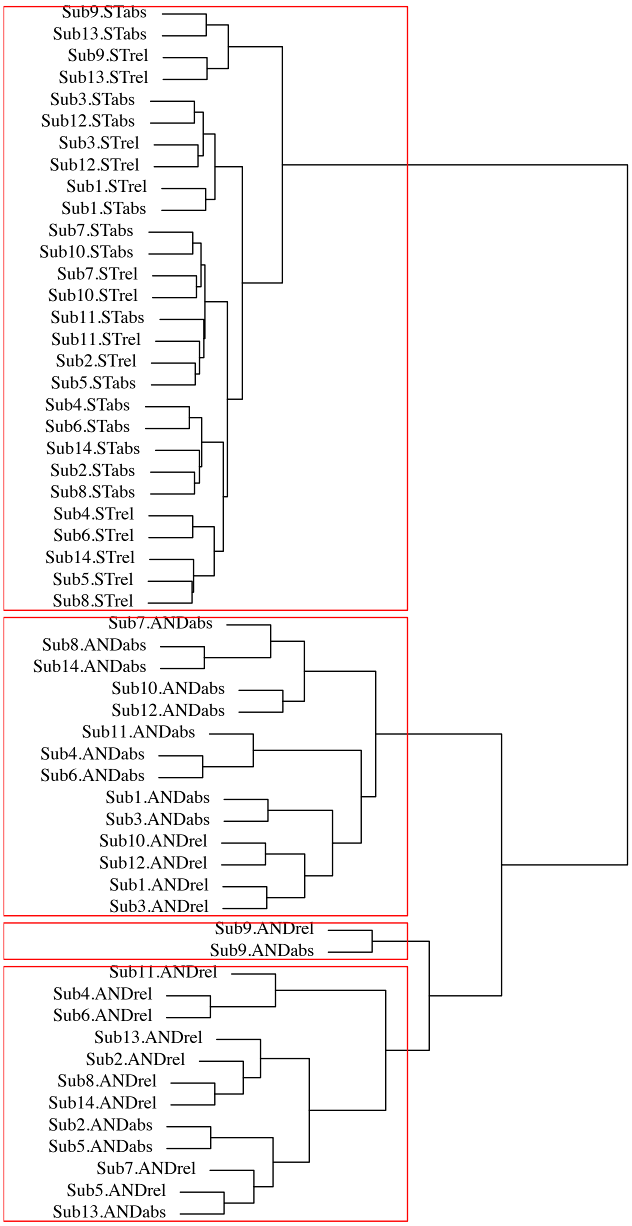

Figure 5: Dendrogram visualization of hierarchical clustering on 20D fPCA score vector space. The red boundaries indicate the cut tree segmentation into group groupings.

through CCF space to get to that same trained end state. But this analysis also revealed that multiple ways of measuring capacity (absolute and relative) were needed to identify these learning patterns.

An alternative to summary statistics is comparing parameters of a fitted model (cf. Eidels, Donkin, Brown, & Heathcote, 2010). The downside to the model fitting approach is that it relies on many more assumptions about how RTs are generated that are ancillary to analyzing the effect of increasing information sources (and hence the degree of configural learning). Ancillary assumptions are necessary in most approaches, including the present analyses (e.g., Euclidean distances metrics). However, measurement assumptions are far less constraining with respect to the potential underlying cognitive processes than direct assumptions about the RT distributions. Clustering and other statistical learning methods applied to the full functional CCF data enables principled, quantified individual differences analysis with minimal assumptions about the best parametric model for capturing the underlying cognitive processes.

Acknowledgments

This work was supported by AFOSR Grant FA9550-13-1-0087 to J. W. H. and AFOSR LRIR to L. M. B. Distribution A. Approved for public release; distribution unlimited. 88ABW Cleared 03/11/2015; 88ABW-2015-0982.

References

Blaha, L. M. (2010). *A dynamic Hebbian-style model of configural learning* (Unpublished doctoral dissertation). Indiana University, Bloomington, Indiana.

Blaha, L. M., Busey, T. A., & Townsend, J. T. (2009, August). An lda approach to the neural correlates of configural learning. In N. A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st annual conference of the cognitive science society.* Austin, TX: Cognitive Science Society.

Burns, D. M., Houpt, J. W., Townsend, J. T., & Endres, M. J. (2013). Functional principal components analysis of workload capacity functions. *Behavior Research Methods*, *45*, 1048-1057.

Eidels, A., Donkin, C., Brown, S. D., & Heathcote, A. (2010). Converging measures of workload capacity. *Psychonomic Bulletin & Review*, *17*, 763-771.

Goldstone, R. L. (2000). Unitization during category learning. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 86-112.

Houpt, J. W., Blaha, L. M., McIntire, J. P., Havig, P. R., & Townsend, J. T. (2014). Systems Factorial Technology with R. *Behavior Research Methods*, *46*, 307-330. doi: 10.3758/s13248-013-0377-3

Houpt, J. W., & Townsend, J. T. (2012). Statistical measures for workload capacity analysis. *Journal of Mathematical Psychology*, *56*, 341-355.

Ramsay, J. O., & Silverman, B. W. (2005). *Functional data analysis* (2nd ed.). New York: Springer.

Townsend, J. T., & Wenger, M. J. (2004). A theory of interactive parallel processing: New capacity measures and predictions for a response time inequality series. *Psychological Review*, *111*, 1003-1035.

Ward, J. H., Jr. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, *58*, 236-244.