Measuring Human Performance on Clustering Problems: Some Potential Objective Criteria and Experimental Research Opportunities

Michael J. Brusco

Abstract

The study of human performance on discrete optimization problems has a considerable history that spans various disciplines. The purpose of this paper is to outline a program of study for the measurement of human performance on discrete optimization problems related to clustering of points in the two-dimensional plane. I describe possible objective criteria for clustering problems, the measurement of agreement of solutions produced by subjects, and categories of experiments for investigating human performance on clustering problems. To facilitate future experimental testing of human subjects on clustering problems, optimal partitions were obtained for 233 two-dimensional clustering problems ranging in size from 10 to 70 points. For each test problem, an optimal solution was obtained for each of three objective criteria: (a) maximizing partition split, (b) minimizing partition diameter, and (c) minimizing within-cluster sums of squares, and similarity of the solutions among these criteria has been computed.

Introduction

Parker and Rardin (1988, chapter 1) characterize *discrete optimization* as a particular class of problems within the much larger field of combinatorics. The defining principle of discrete optimization is the minimization or maximization of some criterion measure over a finite set of mutually exclusive alternatives. There are many relevant discrete optimization problems, and such problems can vary significantly with respect to their computational tractability. A partial list of some of the most familiar discrete optimization problems is as follows: minimum spanning tree, shortest-route, traveling salesperson problem, graph coloring, *p*-median problem, set-covering problem, knapsack problem, bin-packing problem, and quadratic assignment problem. These problems have many important applications in areas such as facility location, vehicle routing, electrical circuitry, assembly line design, telecommunications network architecture, and the analysis of psychological data.

Department of Marketing, College of Business, Florida State University, Tallahassee, FL 32306-1110

Telephone: 850.644.6512; Fax: 850.644.4098; Email: mbrusco@cob.fsu.edu

http://dx.doi.org/10.7771/1932-6246.1001

The study of human performance with respect to discrete optimization tasks spans several problems and academic disciplines. During the 1970s and 1980s, operations research and management science specialists were particularly interested in comparing the performances of human subjects to computer algorithms on a type of quadratic assignment problem related to the location of departments within a facility (Block, 1977; Coleman, 1977; Herroelen & Van Gils, 1985; Scriabin & Vergin, 1975; Trybus & Hopkins, 1980). Subjects were provided with information regarding the flow (or interaction) between each pair of *q* departments, which enabled them to place pairs of departments with high interactions close to one another in the layout. Subjects were also typically provided with a grid consisting of *q* possible locations. Each of *q* departments were to be placed in exactly one of the *q* locations in the grid, which yields a total finite solution space of *q*! possible arrangements. The criterion measure for this discrete optimization problem was to minimize overall weighted flow distance. The emphasis in these studies was primarily on identifying data characteristics that enabled humans to obtain solutions that were as good or better than computer implementations of heuristic algorithms. Understanding the cognitive processes that the human subjects used to develop their solutions was typically of lesser importance.

More recently, considerable research effort has been devoted to human performance on the two-dimensional Euclidean traveling salesperson problem in the field of experimental psychology (Chronicle, MacGregor, Ormerod, & Burr, 2006; Graham, Joshi, & Pizlo, 2000; Lee & Vickers, 2000; MacGregor & Ormerod, 1996, 2000; MacGregor, Ormerod, & Chronicle, 1999, 2000; Ormerod & Chronicle, 1999; van Rooij, Stege, & Schactman, 2003; Vickers, Butavicius, Lee, & Medvedev, 2001; Vickers, Lee, Dry, & Hughes, 2003). In this discrete optimization task, subjects are presented with *N* points in the two-dimensional plane. The subjects are asked to produce a tour that involves leaving from one of the points and sequentially visiting (only once) each of the other points, and then returning to the original point of departure. The feasible solution space consists of $(N–1)!/2$ possible sequences and the criterion measure is to minimize total Euclidean distance traveled. MacGregor et al. (1999) recognized that the justification for investigating human performance on these problems transcends the fundamental interest in human cognitive ability. They observed that a more general study of spatial cognition could be facilitated by studying combinatorial optimization problems. Similarly, Vickers et al. (2001) noted that combinatorial optimization problems could be used to investigate broader, more general issues regarding how the brain constructs models of its environment. These authors further noted the potential use of combinatorial optimization problems in neurophysiological tests.

The investigation of human clustering of points in the two-dimensional plane would seem to be a natural extension of recent research related to the two-dimensional traveling salesperson problem. Instead of constructing a single tour that connects all points, the subjects would partition the points into a collection of clusters or groups. The clustering of points by subjects is not unprecedented in the experimental psychology literature. For example, there is a well-established history of research devoted to the counting or enumeration of points in the two-dimensional plane, and clustering of points to facilitate enumeration is an important component of this research stream (Atkinson, Cambell, & Francis, 1976; Beckwith & Restle, 1966; Klahr, 1973; van Oeffelen & Vos, 1982).

The principal focus of this paper is to link human clustering of points in the two-dimensional plane to relevant partitioning optimization problems. This linkage should facilitate a better understanding of how humans perform on different clustering tasks, as well as what type of criterion individuals naturally employ in a grouping task. The next section of this paper provides a quantitative description of some of the most common partitioning optimization criteria, as well as the advantages and disadvantages of the respective criteria. This is followed by a section that outlines some alternatives for conducting experiments pertaining to human performance on clustering problems. For example, the number of clusters and/or the partitioning criterion could be prespecified by the experimenter, or left to the discretion of the subjects. In either case, the experimenter could employ well-grounded techniques for measuring the agreement of partitions produced by subjects, as well as the agreement between a subject's partition and an optimal partition. The paper concludes with a brief summary.

Clustering Problems

Notation and Assumptions

I limit coverage of clustering problems to situations involving a set of points, $C = \{i = 1, 2, ..., N\}$, in the two-dimensional plane. Each of the N points in the plane is defined by the coordinate pair, (x_i, y_i), for all $1 \leq i \leq N$. It is helpful to define an $N \times N$ matrix of squared Euclidean distances between pairs of points, **D**, where elements of **D** are defined as: $d_{ij} = d_{ij} = (x_i - x_j)^2 + (y_i - y_j)^2$, for $1 \leq i < j \leq N$ and $d_{ii} = 0$ for $1 \leq i \leq N$. The number of clusters is denoted as K, and the subset of points assigned to cluster k is defined as C_k for $1 \leq k \leq K$. Together, the clusters $(C_1, ..., C_K)$ are assumed to define a *partition*, P_K, of the points in C, which implies that the clusters are nonempty ($C_k \neq \varnothing$ for $1 \leq k < l \leq K$), mutually exclusive ($C_k \cap C_l = \varnothing$ for $1 \leq k < l \leq K$), and exhaustive ($C_1 \cup C_2 \cup ..., \cup C_K = C$).

In many applications, the desirable properties of a partition are that the clusters are *homogeneous* and *well-separated*. A homogeneous cluster consists of points that are close to one another. Two clusters are well-separated if there are no points in the first cluster that are close to any point in the second cluster. There are a host of objective criteria that can be used to obtain partitions with clusters that are well-separated and/or homogeneous. I will present three of the most well-known criteria: (a) maximizing partition split, (b) minimizing partition diameter, and (c) minimizing within-cluster sums of squares.

Maximizing Partition Split

The split between clusters C_k and C_l, which is obtained as $split(C_k,C_l) = \min_{l\in C_k, j\in Cl}\{d_j\}$, is the smallest distance from any point in C_k to any point in C_l. The split of the partition is the minimum of the splits between all pairs of clusters, $split(P_K) = \min_{1\le k<l\le K}\{split(C_k,C_l)\}$. The objective of finding a partition, P_K, that maximizes partition split is designed to produce clusters that are well-separated. The problem of finding a K-cluster partition that maximizes partition split is closely related to the problem of finding a *minimum spanning tree* for the points in the two-dimensional plane (see Hubert, 1974a for an especially thorough discussion of spanning trees in cluster analysis). A minimum spanning tree interconnects all points in the plane in minimum total distance. Once the minimum spanning tree is obtained, the maximum split partition is easily produced by breaking the K–1 longest links in the tree. Finding a minimum spanning tree is, in itself, an interesting optimization problem, and Vickers, Mayo, Heitmann, Lee, and Hughes (2004) recently investigated human performance on finding minimum spanning trees. Because the construction of the minimum spanning tree can be done in polynomial time (e.g., Kruskal, 1956), the maximum split optimization problem is a relatively straightforward clustering problem.

Minimizing Partition Diameter

The diameter of cluster C_k is the maximum distance between any pair of points in that cluster, $diameter(C_k) = \max_{(i<j)\in C_k}\{d_j\}$. The diameter of the partition is the maximum of the cluster diameters, $diameter(P_K) = \max_{1\le k\le K}\{diameter(C_k)\}$. The objective of finding a partition, P_K, that minimizes partition diameter is designed to produce clusters that are compact and homogeneous. For the special case of $K = 2$, the problem of minimizing partition diameter can be solved in polynomial time using an algorithm designed by Rao (1971); however, Brücker (1978) and Hansen and Delattre (1978) showed that minimum diameter partitioning is NP-hard for $K \ge 3$. Fortunately, branch-and-bound methods can often facilitate optimal solution of minimum diameter partitioning problems with large N and K (Brusco & Cradit, 2004; Brusco & Stahl, 2005, chapter 3; Hansen & Delattre, 1978).

Minimizing Within-Cluster Sums of Squares

Perhaps the most popular partitioning criterion is the minimization of the within-cluster sums of squared deviations from the cluster centroids. This criterion is most typically associated with the well-known K-means clustering algorithms (Forgy, 1965; MacQueen, 1967; Hartigan & Wong, 1979), and an excellent review of this literature was recently provided by Steinley (2006). For the two-dimensional case, the within-cluster sums of squares (WCSS) criterion for a partition, P_K, is computed as follows:

$$WCSS(P_K) = \sum_{k=1}^{K}\sum_{i\in C_k}\left[(x_i - \bar{x}_k)^2 + (y_i - \bar{y}_k)^2\right], \tag{1}$$

where $\bar{x}_k = \frac{\sum_{i \in C_k} x_i}{n_k}$ and $\bar{y}_k = \frac{\sum_{i \in C_k} y_i}{n_k}$ are the means of the *x* and *y* coordinates in cluster *k* (for $1 \leq k \leq K$), respectively, and $n_k = |C_k|$ is the number of points assigned to cluster *k* (for $1 \leq k \leq K$). Using Huygens's theorem (see Edwards & Cavalli-Sforza, 1965 or Späth, 1980, chapter 3), it is possible to represent $WCSS(P_K)$, using matrix **D** as follows:

$$WCSS(P_K) = \sum_{k=1}^{K} \left[\frac{\sum_{(i<j) \in C_k} d_{ij}}{n_k} \right]. \quad (2)$$

Finding a partition that minimizes $WCSS(P_K)$ is an NP-hard optimization problem (Brücker, 1978). Dynamic programming methods (Hubert, Arabie, & Meulman, 2001; van Os & Meulman, 2004) can produce optimal solutions for problems with roughly 25 to 30 points, and branch-and-bound procedures can produce optimal solutions for much larger problems (Brusco, 2006; Brusco & Stahl, 2005, chapter 5; Koontz, Narendra, & Fukunaga, 1975). The effectiveness of the latter class of procedures depends not only on the number of points, but also on the number of clusters and the separation between clusters. For randomly generated points in the two-dimensional plane, Brusco successfully solved problems with up to $N = 60$ points and $K = 6$ clusters in a reasonable amount of time.

A Numerical Example

To demonstrate the partitioning criteria, I use a small numerical example from Brusco and Stahl (2005, p. 67). The data consist of $N = 6$ points in the two-dimensional plane, and the coordinates of the points are: ($x_1 = 7, y_1 = 2$), ($x_2 = 2, y_2 = 4$), ($x_3 = 5, y_3 = 4$), ($x_4 = 4, y_4 = 6$), ($x_5 = 3, y_5 = 1$), ($x_6 = 8, y_6 = 4$). Optimal two-cluster partitions for these data for the criteria of partition split, partition diameter, and WCSS are displayed in Figure 1. The values of $split(P_K)$, $diameter(P_K)$, and $WCSS(P_K)$ for each partition in Figure 1 are reported in Table 1.

Table 1. Splits, diameters, and within-cluster sums of squares for the three partitions in Figure 1.

Optimization Criterion	Optimal Partition (P_K)	Panel in Figure 1	$split(P_K)$	$diameter(P_K)$	$WCSS(P_K)$
Maximize	{1,2,3,4,6} {5}	Top	10	36	30.80
Minimize	{1,3,4,6} {2,5}	Middle	8	25	23.00
Minimize	{1,6} {2,3,4,5}	Bottom	8	26	20.25

The partition in the top panel of Table 1 places five points in one cluster {1,2,3,4,6} and leaves point 5 in its own individual cluster. The split between these two clusters is the distance between point 5 and its closest neighbor among the remaining points, which is point 2 at a distance of 10. Thus, $split(P_K) = 10$ for the partition in the top panel, and

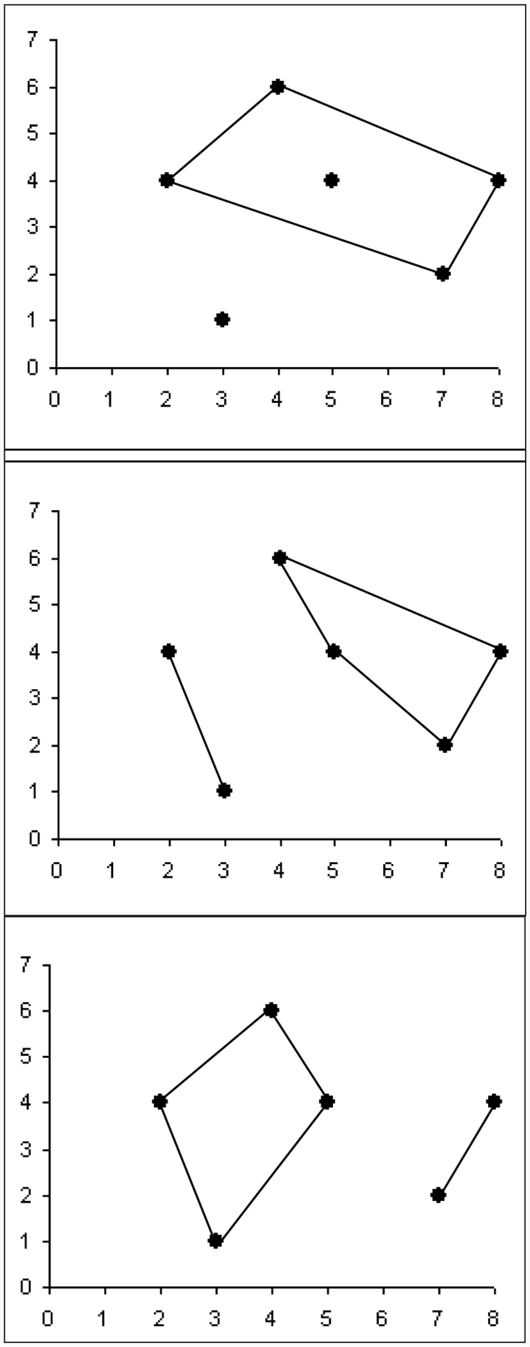

Figure 1. Optimal two-cluster partitions for criteria of: maximum partition split (top panel), minimum partition diameter (middle panel), and minimum within-cluster sum of squares (bottom panel).

this is the maximum split among all possible two-cluster partitions. Although the clusters are well-separated, they are not compact, as the two most distant points in the data set (points 2 and 6) are in the same cluster. The relatively poor values of $diameter(P_K) = 36$ and $WCSS(P_K) = 30.80$ reflect the lack of homogeneity for the partition in the top panel in Figure 1.

The partitions in the middle and bottom panels of Figure 1 exhibit much greater cluster homogeneity than the partition in the top panel. The cluster diameters for the middle-panel partition are $d_{25} = 10$ for the cluster {2,5} and $d_{14} = 25$ for the cluster {1,3,4,6}. Therefore, the partition diameter is max(10, 25) = 25, which is the minimum partition diameter. Although the partition in the bottom panel is markedly different, its partition diameter is only slightly larger than the optimal value. The cluster diameters for the bottom-panel partition are $d_{16} = 5$ for the cluster {1,6} and $d_{45} = 26$ for the cluster {2,3,4,5}, resulting in a partition diameter of max(5, 26) = 26. Although the middle-panel partition has a slightly smaller value of $diameter(P_K)$ than the bottom-panel partition, the middle-panel partition's $WCSS(P_K) = 23.00$ is somewhat larger than the optimal value of $WCSS(P_K) = 20.25$ corresponding to the bottom-panel partition. The middle-panel and bottom-panel partitions in Table 1 both yield $split(P_K) = 8$.

A Larger Numerical Example

The preceding example is for a rather small synthetic data set, so I will provide a second example for a well-studied data set originally reported by Späth (1980, p. 43), which corresponds to the coordinates for 22 German cities (see Figure 2). Four-cluster partitions for these data were obtained for the maximum split, minimum diameter, and minimum WCSS criteria. The minimum spanning tree for the German cities data is displayed in Figure 3. The three dashed edges in the spanning tree are those edges that would be deleted to produce a maximum split, four-cluster partition.

Brusco and Stahl (2005, Chapter 5) obtained minimum WCSS partitions for the German cities data for $2 \leq K \leq 8$. Figure 4 provides the optimal four-cluster partition, which is represented by solid boundaries. The cities assigned to the four labeled clusters are C1 {Kiel, Lübeck, Hamburg, Bremen, Braunschweig}, C2 {Aachen, Köln, Essen, Münster, Bielefeld, Kassel}, C3 {Saarbrücken, Mannheim, Freiburg, Karlsruhe}, and C4 {Würzburg,

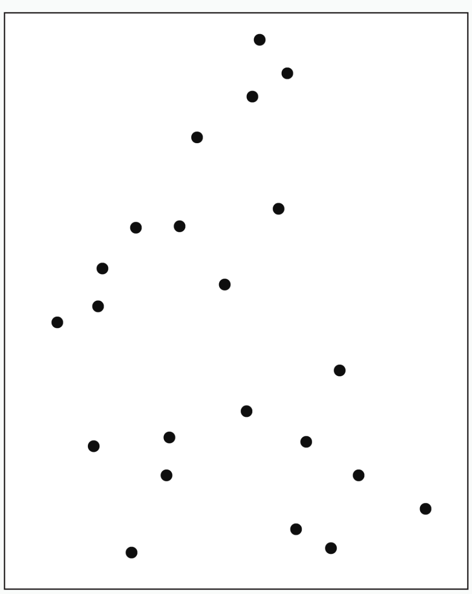

Figure 2. A plot of Späth's (1980, p. 43) coordinates for 22 German cities.

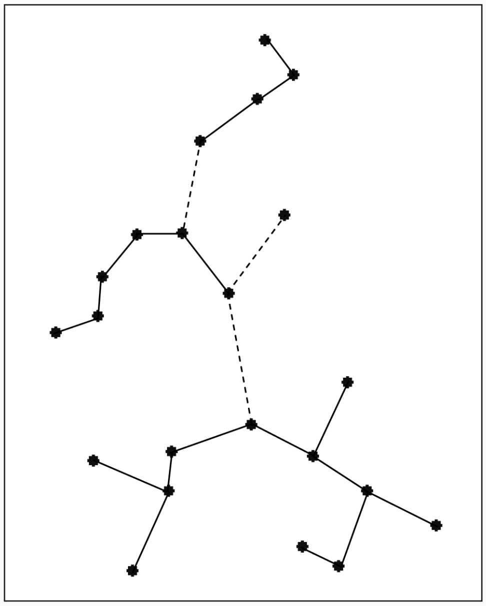

Figure 3. Minimum spanning tree for Späth's (1980, p. 43) data for 22 German cities. The 3 dashed lines are the ones that would be broken to produce a maximum split four-cluster partition.

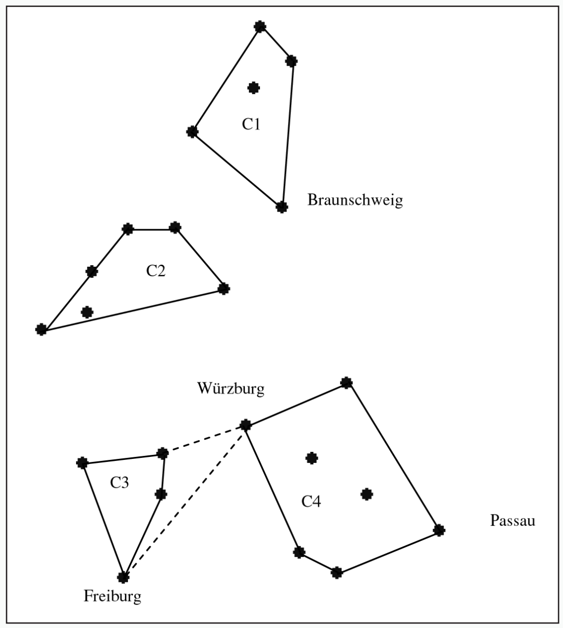

Figure 4. Optimal four-cluster partitions for Späth's (1980, p. 43) data for 22 German cities. The solid lines show the minimum within-cluster sum of squares partition. The dashed lines indicate that moving Würzburg from C4 to C3 produces a minimum-diameter partition. Joining C4 and C3 and moving Braunschweig to its own individual cluster will produce the maximum split partition in accordance with the minimum spanning tree in Figure 2.

Augsburg, München, Nürnberg, Regensburg, Hof, Passau}. A minimum diameter four-cluster partition can be obtained from the minimum WCSS partition by moving Würzburg from C4 to C3, as shown by the dashed lines in Figure 4. This move reduces the partition diameter because the distance from Würzburg to Passau is greater than the distance from Würzburg to Freiburg. Whereas the optimal partitions for the two homogeneity criteria (diameter and WCSS) are nearly identical, they are markedly different from the optimal partition for split suggested in Figure 3. To obtain the maximum split four-cluster partition from the minimum WCSS (or minimum diameter) partition, clusters C3 and C4 would be merged, and a new, singleton cluster would be produced by removing Braunschweig from C1.

Critiquing Alternative Clustering Criteria

The examples in the preceding subsections show that, even for a small data set, three straightforward criteria can yield three different optimal partitions. For this reason, some discussion of the advantages and disadvantages of each criterion for measuring human performance is appropriate. Partition split and partition diameter are classic criteria with important links to hierarchical clustering as well as partitioning (Hubert, 1974b; Johnson, 1967). These two criteria also have the advantage of monotone invariance. In other words, any order-preserving transformation of matrix **D** will not affect the optimal partition associated with the split and diameter criteria. If I repeated the numerical examples in the previous subsections after converting the squared Euclidean distances in **D** to Euclidean distances by taking the square root of each entry in **D**, the optimal partitions for split and diameter would remain unchanged. The within-cluster sums of squares criterion, as well as many related criteria that use sums of deviations, is not necessarily invariant to order-preserving transformations.

A possible disadvantage of the split criterion is a tendency to produce one or more singleton clusters (i.e., clusters with only one point) in the partition. Singleton clusters can also occur when minimizing diameter is the objective, however, the principal disadvantage of this criterion is the potential for a large number of alternative optimal partitions. In applied cluster analyses, problems associated with alternative optima can be addressed by enumerating all optimal partitions for modestly sized data sets (Guénoche, 1993), or by tie-breaking based on some secondary criterion (Brusco & Cradit, 2004, 2005). However, in an experimental study of human performance, 10 subjects could produce 10 different partitions and each partition could be of minimum diameter. The importance of measuring agreement of partitions across subjects would, therefore, be at least as important as measuring proximity to the optimal criterion value.

The within-cluster sums of squares criterion is less susceptible to problems associated with alternative optima; however, it too has some potential drawbacks. It is well known that the criterion tends to produce spherical clusters of approximately the same size. In some clustering applications, clusters might assume more of an elliptical shape, perhaps with different spatial orientations and sizes. Alternative criteria for various shapes and orientations, which are principally based on determinants of within-cluster sums of squares matrices and/or between-cluster sums of squares matrices have been proposed by a number of authors (Banfield & Raftery, 1993; Friedman & Rubin, 1967; Maronna & Jacovkis, 1974; Marriott, 1982; Scott & Symons, 1971; Symons, 1981; Windham, 1987).

Human Performance Experiments in Cluster Analysis

The development of experiments to study human performance on clustering problems can employ a number of different design strategies. Stimuli for experiments could be produced by randomly generating points in a two-dimensional plane, or by using previously published two-dimensional data sets from the clustering and related literature bases. Given a selected set of stimuli, I consider three possible categories for experiments, which progressively allocate more freedom to the subject in producing a clustering solution. The three categories are: (a) fixed *K* with a specific clustering criterion, (b) fixed *K* without a specific clustering criterion, and (c) flexible *K* without a specific clustering criterion.

Fixed K, Specific Criterion

In this type of experiment, subjects would be instructed to produce *K* clusters of the points using an objective criterion defined by the user (e.g., maximize split, minimize diameter, etc.). The communication of the task to subjects could potentially have considerable bearing on the quality of solutions produced. Some subjects might operate under a tacit assumption that clusters should be of roughly equal size, even for problems that have one or more small clusters or possibly even singleton clusters. Thus, the description of the task might not need only to describe the criterion, but also the concept of a partition and the admissible types of solutions. Performance comparisons could take at least three forms: (a) percentage deviation between the objective criterion associated with subject partitions and the optimal objective criterion value, (b) agreement between the optimal partition and the partitions provided by subjects, and (c) agreement among partitions provided across subjects.

For measuring partition agreement in clustering experiments, I recommend Hubert and Arabie's (1985) adjusted Rand index (ARI) as a measurement of agreement between two partitions. The ARI between two partitions, P^1 and P^2, is well-recognized and widely used in the classification literature (see Steinley, 2004 for an evaluation and review of the index). Table 2 facilitates a description of the ARI, which considers all $N(N–1)/2$ pairs of points (point pairs) in each of the two partitions. With respect to a given point pair (i, j), agreement between the two partitions occurs when points i and j are in the same cluster in P^1 and in the same cluster in P^2. Agreement also arises when points i and j are in different clusters in P^1 and in different clusters in P^2. Thus, disagreement only occurs when points i and j are in the same cluster in one partition, but in different clusters in the other partition. The ARI is computed as follows:

$$\mathrm{ARI} = \frac{(N(N-1)/2)\,(\tau_1+\tau_2) - [(\tau_1+\tau_3)(\tau_1+\tau_4) + (\tau_2+\tau_3)(\tau_1+\tau_2)]}{(N(N-1)/2) - [(\tau_1+\tau_3)(\tau_1+\tau_4) + (\tau_2+\tau_3)(\tau_1+\tau_2)]}. \quad (3)$$

where τ_1 is the number of point pairs in same cluster for both P^1 and P^2, τ_2 is the number of point pairs in different clusters for both P^1 and P^2, τ_3 is the number of point pairs in the same cluster in P^1 but different clusters for P^2, and τ_4 is the number of point pairs in the same cluster in P^2 but different clusters for P^1. An ARI = 1, which occurs when $\tau_3 = \tau_4 = 0$, indicates perfect agreement between P^1 and P^2, whereas ARI near zero suggests chance agreement.

Table 2. Hubert and Arabie's (1985) adjusted Rand index.

		Partition 2		
		Number of Point Pairs in Same Cluster	Number of Point Pairs in Different Clusters	Totals
Partition 1	Number of Point Pairs in Same Cluster	τ_1	τ_3	$\tau_1 + \tau_3$
	Number of Point Pairs in Different Clusters	τ_4	τ_2	$\tau_2 + \tau_4$
	Totals	$\tau_1 + \tau_4$	$\tau_2 + \tau_3$	$N(N-1)/2$

To demonstrate the ARI, I return to the example in Figure 1. The partition in the middle panel of Figure 1 has the same split and nearly the same diameter as the partition in the bottom panel; however, the agreement between these two partitions as measured by the ARI is only −.17. In contrast, the middle-panel partition has a different split and much different diameter than the partition in the top panel; however, the ARI between these two partitions is .35. The key here is that there is not necessarily a one-to-one relationship between ARI and any given objective criterion value.

For a fixed *K* experiment with a specific criterion, one possible avenue for investigation is relative subject performance as a function of *N* and *K*. For example, consider an experiment where the stimuli are produced by randomly generating 20, 40, or 80 points in the plane. For a selected criterion, subjects could be asked to produce two-cluster, four-cluster, and eight-cluster solutions for each of these stimuli. This results in a two-factor design with three levels for the number of points, *N*, and three levels for the number of clusters, *K*.

It could also be valuable to compare subject performances across different objective criteria. That is, for the same two-factor design, an experimenter could have subjects produce solutions given the criterion of maximum split, as well as solutions given the criterion of minimum diameter. Important research questions under this context might include: (a) do subjects exhibit better performance (relative to the optimum or simple heuristic procedures) for split or diameter?, and (b) is there greater agreement among subject partitions for the criterion of split, or for the criterion of diameter?

Fixed K, *No Specific Criterion*

The second category of experiments removes one of the constraints on the subjects by not imposing a specific clustering criterion. This category could be especially important for understanding how humans would naturally tend to cluster points. Unfortunately, experiments in this category present an additional obstacle with respect to communication of the task. In addition to understanding the concept of partition and feasible solutions for the task, the context of the question could affect the types of solutions that subjects produce. For example, telling the subjects that the points in the plane are the geographic coordinates of towns might induce results that are different than those produced by subjects who are told that the points are birth and death rates for various countries.

Despite the challenge of presenting an appropriate task description, experiments that do not impose a specific criterion offer a potentially valuable contribution. For example, two plausible questions that could be explored in this category are: 1) Do human subjects tend to focus more heavily on separation between the clusters they produce, or homogeneity within the clusters they produce? and 2) Do subjects favor the maximin criterion (split), the minimax criterion (diameter), or the minisum criterion (WCSS) in the solutions they generate? These questions could be assessed partially by comparing subject partitions to optimal partitions for split, diameter, and WCSS. As a simple example, suppose that the six-point stimulus used in Figure 1 was provided with a simple instruction to subjects to partition the points into two clusters. Which panel in Figure 1 would be more frequently provided by subjects? Is there a partition other than those in Figure 1 that would be more popular? What is the agreement of partitions across subjects as measured by ARI?

Flexible K, *No Specific Criterion*

This least restrictive experimental category would simply ask subjects to produce a partition of points without any criterion or designated number of clusters. In addition to experimental questions along the lines of those posited for fixed K (no specific criterion), another interesting issue arises with respect to the number of clusters used by subjects. What is the variability of the number of clusters used by subjects? As the number of clusters increases, do subjects tend to focus more on partition diameter or more on split, or are they concerned with both separation and homogeneity in their solutions?

Failure to impose any constraint on K can have serious ramifications for the solutions produced by subjects. For example, one subject might examine a stimulus of points in the two-dimensional plane and conclude that there are no definitive clusters, thus choosing $K = 1$ (all points in one cluster) as the solution. Another subject could consider the same stimulus and place each point in its own cluster ($K = N$) because of a failure to observe any patterning in the data points. The remaining subjects could produce solutions that span the range of $1 \leq K \leq N$ clusters, which would make it difficult to draw any meaningful findings from the results. One possible remedy for this type of problem is to place a constraint on the permissible range for K; however, as this range is tightened, the problem approaches the fixed K problem category described above.

Computational Results for Selected Test Problems

I have compiled a database of optimal partitions for a number of two-dimensional test problems from the human performance and clustering literature. Table 3 provides a description of the selected test problems, including the original source, the number of data points, and the range of K for which optimal solutions were obtained. Some of the data sets in Table 3 were synthetically constructed (MacGregor & Ormerod, 1996), others represent the location of towns (Späth, 1980, pp. 43, 80), and one corresponds to birth and death rates in countries (Hartigan, 1975, chapter 11). There are 233 unique test problems, and optimal partitions for the split, diameter, and WCSS criteria were obtained for each

Table 3. Test problems selected from the literature.

Problem label	Data source	Number of points, n	Range of clusters	Total number of test problems
DFJ_10	Dantzig, Fulkerson, & Johnson (1959)	10	$2 \leq K \leq 9$	8
MO6_10	MacGregor & Ormerod (1996, p. 539)	10	$2 \leq K \leq 9$	8
MO5_10	MacGregor & Ormerod (1996, p. 539)	10	$2 \leq K \leq 9$	8
MO4_10	MacGregor & Ormerod (1996, p. 539)	10	$2 \leq K \leq 9$	8
MO3_10	MacGregor & Ormerod (1996, p. 539)	10	$2 \leq K \leq 9$	8
MO2_10	MacGregor & Ormerod (1996, p. 539)	10	$2 \leq K \leq 9$	8
MO1_10	MacGregor & Ormerod (1996, p. 539)	10	$2 \leq K \leq 9$	8
MO16_20	MacGregor & Ormerod (1996, p. 539)	20	$2 \leq K \leq 19$	18
MO14_20	MacGregor & Ormerod (1996, p. 539)	20	$2 \leq K \leq 19$	18
MO12_20	MacGregor & Ormerod (1996, p. 539)	20	$2 \leq K \leq 19$	18
MO10_20	MacGregor & Ormerod (1996, p. 539)	20	$2 \leq K \leq 19$	18
MO8_20	MacGregor & Ormerod (1996, p. 539)	20	$2 \leq K \leq 19$	18
MO6_20	MacGregor & Ormerod (1996, p. 539)	20	$2 \leq K \leq 19$	18
MO4_20	MacGregor & Ormerod (1996, p. 539)	20	$2 \leq K \leq 19$	18
S_22	Späth (1980, p. 43)	22	$2 \leq K \leq 21$	20
KFM_48*	Krolak, Felts, & Marble (1971, p. 332)	48	$2 \leq K \leq 18$	17
S_59	Späth (1980, p. 80)	59	$2 \leq K \leq 8$	7
H_70	Hartigan (1975, Ch. 11)	70	$2 \leq K \leq 8$	7

*MacGregor et al. (1999) randomly extracted 48 of the 100 points in the Krolak et al. (1971) test problem.

test problem. Optimal partitions for diameter and WCSS were obtained using the branch-and-bound programs described by Brusco and Stahl (2005, chapters 3, 5), and a Fortran program was written to obtain the maximum split partition from a spanning tree. These programs can be obtained from the website <http://garnet.acns.fsu.edu/~mbrusco>.

For smaller test problems, optimal solutions were obtained for $2 \le K \le N-1$. However, for the larger test problems ($N \ge 48$), this was not computationally feasible for the WCSS criterion. Therefore, optimal solutions were limited to smaller values of K for the larger test problems. I computed the ARI between each pair of optimal partitions (split vs. diameter, split vs. WCSS, and diameter vs. WCSS). These results should be used only as a guideline for relative agreement of optimal partitions among the criteria because they do not account for alternative optimal partitions for each criterion. Across the 233 test problems, the average ARI values were .49, .52, and .70 for the split vs. diameter, split vs. WCSS, and diameter vs. WCSS comparisons, respectively.

There are many test problems where the agreement among the optimal partitions for the three criteria is mediocre or poor, and these could provide exceptional candidates for human performance experiments. For example, consider the 48-point data set from Krolak et al. (1971, p. 332), which is displayed in Figure 5a. Five-cluster partitions of this data set for split, diameter, and WCSS are displayed in Figures 5b, 5c, and 5d, respectively. The ARI values are .56, .58, and .64 for the split vs. diameter, split vs. WCSS, and diameter vs. WCSS comparisons, respectively. Although there are clearly similarities among the partitions shown in Figures 5b through 5d, there are also some marked differences depending on the criterion selected. For example, the maximum split partition in Figure

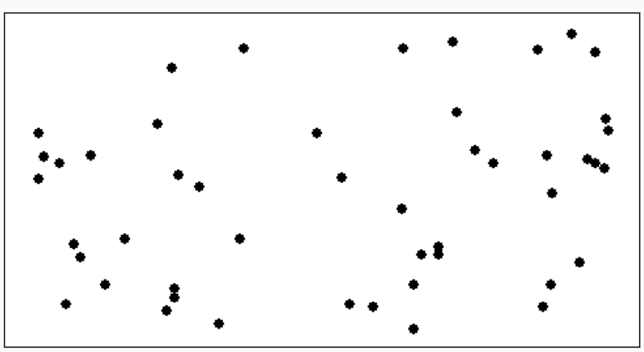

Figure 5a. A plot of 48-points from Krolak et al.'s (1971) data set.

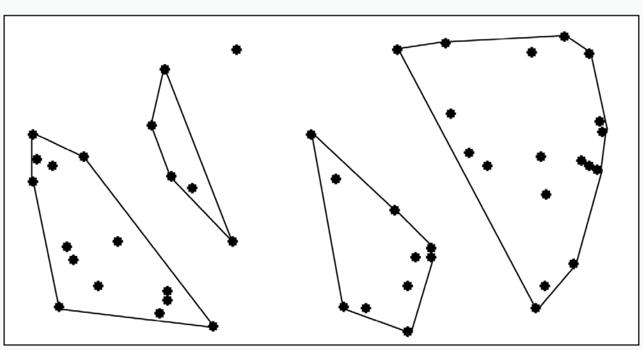

Figure 5b. Maximum split partition for Krolak et al.'s (1971) 48-point data set when using $K = 5$ clusters.

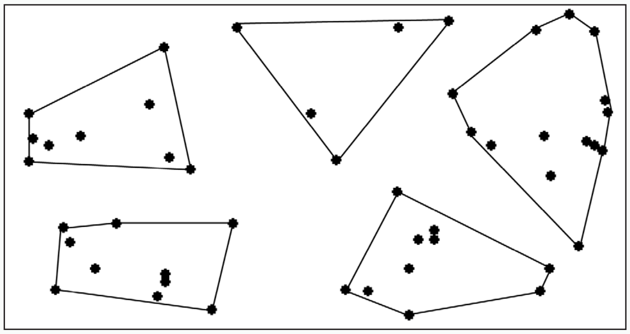

Figure 5c. Minimum diameter partition for Krolak et al.'s (1971) 48-point data set when using $K = 5$ clusters.

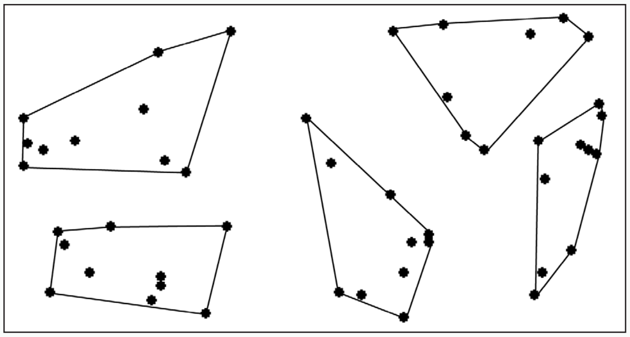

Figure 5d. Minimum WCSS partition for Krolak et al.'s (1971) 48-point data set when using $K = 5$ clusters.

5b has a singleton cluster, as well as a large cluster on the right side of the figure. The large cluster on the right is subdivided in the minimum diameter and minimum WCSS partitions in Figures 5c and 5d; however, the manner in which the cluster is subdivided differs in these two figures.

The 70-point data set from Hartigan (1975, chapter 11) is visually displayed in Figure 6a. Five-cluster partitions of this data set for split, diameter, and WCSS are displayed in Figures 6b, 6c, and 6d, respectively. The ARI values are .17, .04, and .48 for the split vs. diameter, split vs. WCSS, and diameter vs. WCSS comparisons, respectively. There is significant disparity among the partitions shown in Figures 6b through 6d. The maximum split partition in Figure 6b has three singleton clusters, one cluster with two points, and one large cluster with 65 points. The minimum diameter and minimum WCSS partitions in Figures 6b and 6c

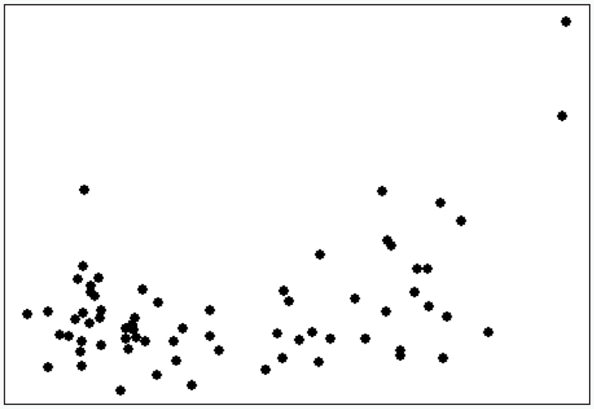

Figure 6a. A plot of 70-points (birth/death rates) from Hartigan (1975) data set.

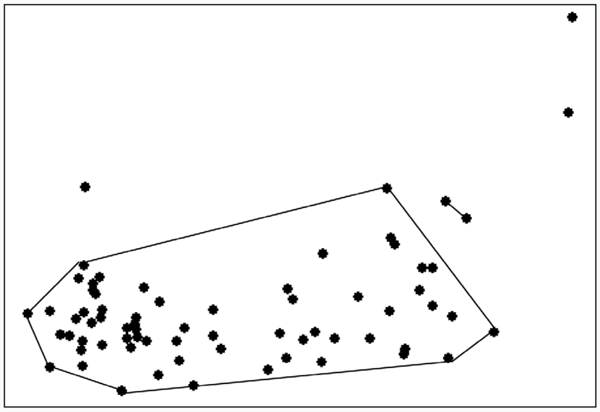

Figure 6b. Maximum split partition for Hartigan's (1975) 70-point data set when using $K = 5$ clusters.

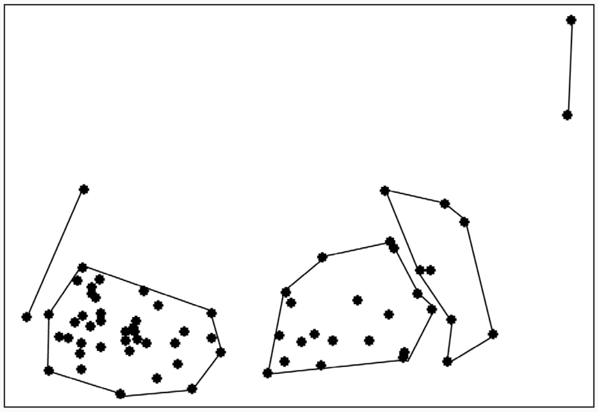

Figure 6c. Minimum diameter partition for Hartigan's (1975) 70-point data set when using $K = 5$ clusters.

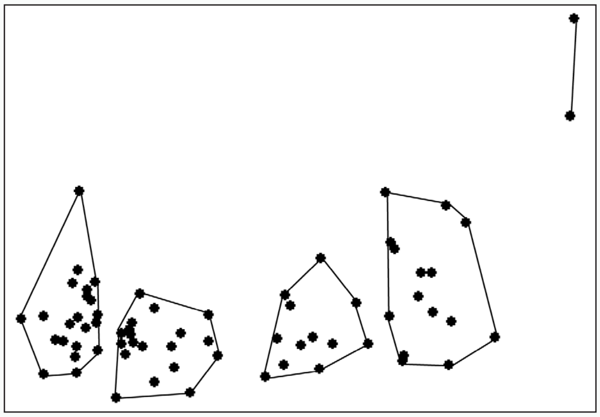

Figure 6d. Minimum WCSS partition for Hartigan's (1975) 70-point data set when using $K = 5$ clusters.

both place the two extreme points in the upper right corner in the same cluster. The WCSS partition, however, has a more equitable balance of points across the remaining clusters.

Discussion

My goal in this paper was to outline some basic types of experiments that could expand research in human performance on optimization problems to areas of cluster analysis. I have couched most of the discussion within the simplest measures of cluster separation (split) and cluster homogeneity (diameter), with some lesser attention to within-cluster sums of squares partitioning because of its esteemed position in the classification literature. The split and diameter criteria were selected because they are especially easy to describe to human subjects. The WCSS criterion can be described with a bit more effort, but it would be virtually impossible to ask the average subject to produce a partition that optimizes some function of the determinant of a within-cluster sums of squares matrix. A possible alternative to the WCSS criterion is the *K*-median model (Klastorin, 1985), which has two advantages. First, the centroids of a *K*-median model correspond to actual points, as opposed to the "virtual" centroids of the WCSS solution that are averages across points in the cluster. Second, relative to WCSS, optimal solutions for two-dimensional *K*-median problems can generally be obtained for much larger *N* and *K*.

I have also attempted to outline, in broad terms, some possible clustering experiments. The principal goal was to identify some of the relevant controls on experiments (the stated objective and value of *K*). I have not proposed models of cognitive and visual processes to offer meaningful hypotheses regarding these types of experiments, nor have I addressed the issue of fitting mathematical models to reflect human performances. For sufficient understanding of the processes that underlie how humans cluster, it might well be necessary to consider more sophisticated, model-based clustering procedures (Banfield & Raftery, 1993). For example, none of the criteria offered in this paper would adequately represent how subjects would cluster a stimulus data set consisting of a few elliptical clusters with various spatial orientations.

There are also opportunities for integrating clustering and traveling salesperson problems in subsequent human performance experiments. For example, one generalization of the traveling salesperson problem requires multiple routes to stem from the same origin. Consider, for example, a delivery vehicle that must make deliveries to 20 warehouses from a storage depot over a period of two days (Saturday and Sunday). The key questions are: (1) Which warehouses should receive deliveries on Saturday, and which ones should receive deliveries on Sunday? and (2) What are the optimal routes on each day? The objective is to minimize total distance traveled over the two-day period. Clearly, this problem involves a clustering problem because of the need to partition the warehouses into Saturday and Sunday deliveries. However, there is also a traveling salesperson problem for each day. How would human subjects solve two-dimensional Euclidean representations of this type of problem? Would they cluster the points first and then seek the routes? Alternatively, would subjects tend to begin by sketching out the routes and use them to help determine the clusters? There are many interesting problems that can have both clustering and routing components.

Author note

I am grateful to the Editor and two anonymous reviewers for helpful comments that led to tremendous improvements in this paper.

References

Atkinson, J., Cambell, F. W., & Francis, M. R. (1976). The dependence of the visual numerosity limit on orientation, colour, and grouping in the stimulus. *Perception*, *5*, 335–342.

Banfield, J. D., & Raftery, A. E. (1993). Model-based Gaussian and non-Gaussian clustering. *Biometrics*, *49*, 803–821.

Beckwith, M., & Restle, F. (1966). Process of enumeration. *Psychological Review*, *73*, 437–444.

Block, T. L. (1977). A note on "Comparison of algorithms and visual based methods for plant layout" by M. Scriabin and R. C. Vergin. *Management Science*, *24*, 235–237.

Brücker, F. (1978). On the complexity of clustering problems. In M. Beckmann & H. P. Kunzi (Eds.), *Optimization and operations research* (pp. 45–54). Heidelberg: Springer-Verlag.

Brusco, M. J. (2006). A repetitive branch-and-bound procedure for minimum within-cluster sums of squares partitioning. *Psychometrika*, *71*, 347–363.

Brusco, M. J., & Cradit, J. D. (2004). Graph coloring, minimum-diameter partitioning, and the analysis of confusion matrices. *Journal of Mathematical Psychology*, *48*, 301–309.

Brusco, M. J., & Cradit, J. D. (2005). Bicriterion methods for partitioning dissimilarity matrices. *British Journal of Mathematical and Statistical Psychology*, *58*, 333–343.

Brusco, M. J., & Stahl, S. (2005). *Branch-and-bound applications in combinatorial data analysis*. New York: Springer.

Chronicle, E. P., MacGregor, J. N., Ormerod, T. C., & Burr, A. (2006). It looks easy! Heuristics for combinatorial optimization problems. *Quarterly Journal of Experimental Psychology, 59,* 783–800.

Coleman, D. R. (1977). Plant layout: Computers versus humans. *Management Science*, *24*, 107–112.

Dantzig, G. B., Fulkerson, D. R., & Johnson, S. M. (1959). On a linear-programming, combinatorial approach to the traveling salesman problem. *Operations Research*, *7*, 58–66.

Edwards, A. W. F., & Cavalli-Sforza, L. L. (1965). A method for cluster analysis. *Biometrics*, *21*, 362–375.

Forgy, E. W. (1965). Cluster analyses of multivariate data: Efficiency versus interpretability of classifications. *Biometrics*, *21*, 768.

Friedman, H. P., & Rubin, J. (1967). On some invariant criteria for grouping data. *Journal of the American Statistical Association*, *62*, 1159–1178.

Graham, S. M., Joshi, A., & Pizlo, Z. (2000). The traveling salesman problem: A hierarchical model. *Memory & Cognition*, *28*, 1191–1204.

Guénoche, A. (1993). Enumération des partitions de diamètre minimum [Enumeration of the partitions of minimum diameter]. *Discrete Mathematics*, *111*, 277–287.

Hansen, P., & Delattre, M. (1978). Complete-link cluster analysis by graph coloring. *Journal of the American Statistical Association*, *73*, 397–403.

Hartigan, J. A. (1975). *Clustering algorithms*. New York: Wiley.

Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS136: A *k*-means clustering program. *Applied Statistics*, *28*, 100–128.

Herroelen, W., & Van Gils, A. (1985). On the use of flow dominance in complexity measures for facility layout problems. *International Journal of Production Research*, *23*, 97–108.

Hubert, L. J. (1974a). Spanning trees and aspects of clustering. *British Journal of Mathematical and Statistical Psychology*, *27*, 14–28.

Hubert, L. J. (1974b). Some applications of graph theory to clustering. *Psychometrika*, *39*, 283–309.

Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification*, *2*, 193–218.

Hubert, L., Arabie, P., & Meulman, J. (2001). *Combinatorial data analysis: Optimization by dynamic programming*. Philadelphia: Society for Industrial and Applied Mathematics.

Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika*, *32*, 241–254.

Klahr, D. (1973). A production system for counting, subitizing, and adding. In W. G. Chase (Ed.), *Visual information processing*. New York: Academic Press.

Klastorin, T. (1985). The *p*-median problem for cluster analysis: A comparative test using the mixture model approach. *Management Science*, *31*, 84–95.

Koontz, W. L. G., Narendra, P. M., & Fukunaga, K. (1975). A branch and bound clustering algorithm. *IEEE Transaction on Computers*, *C-24*, 908–915.

Krolak, P., Felts, W., & Marble, G. (1971). A man-machine approach toward the traveling salesman problem. *Communications of the ACM*, *14*, 327–334.

Kruskal, J. B. (1956). On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical Society*, *7*, 48–50.

Lee, M. D., & Vickers, D. (2000). The importance of the convex hull for human performance on the traveling salesman problem: A comment upon MacGregor and Ormerod (1996). *Perception & Psychophysics*, *62*, 226–228.

MacGregor, J. N., & Ormerod, T. (1996). Human performance on the traveling salesman problem. *Perception & Psychophysics*, *58*, 537–539.

MacGregor, J. N., & Ormerod, T. (2000), Evaluating the importance of the convex hull in solving the Euclidean version of the traveling salesperson problem: Reply to Lee and Vickers (2000). *Perception & Psychophysics*, *62*, 1501–1503.

MacGregor, J. N., Ormerod, T. C., & Chronicle, E. P. (1999). Spatial and contextual factors in the traveling salesperson problem. *Perception*, *28*, 1417–1428.

MacGregor, J. N., Ormerod, T. C., & Chronicle, E. P. (2000). A modeling of human performance on the traveling salesperson problem. *Memory & Cognition*, *28*, 1183–1190.

MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. In L. M. Le Cam & J. Neyman (Eds.) *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. (Vol. 1, 281–297). Berkeley, CA: University of California Press.

Maronna, R., & Jacovkis, P. M. (1974). Multivariate clustering procedures with variable metrics. *Biometrics*, *30*, 499–505.

Marriott, F. H. C. (1982). Optimization methods of cluster analysis. *Biometrika*, *69*, 417–421.

Ormerod, T. C., & Chronicle, E. P. (1999). Global perceptual processes in problem solving: The case of the traveling salesperson. *Perception & Psychophysics*, *61*, 1227–1238.

Parker, R. G., & Rardin R. L. (1988). *Discrete optimization*. San Diego: Academic Press.

Rao, M. R. (1971). Cluster analysis and mathematical programming. *Journal of the American Statistical Association*, *66*, 622–626.

Scott, A. J., & Symons, M. J. (1971). Clustering methods based on likelihood ratio criteria. *Biometrics*, *27*, 387–398.

Scriabin, M., & Vergin, R. C. (1975). Comparison of computer algorithms and visual based methods for plant layout. *Management Science*, *22*, 172–181.

Späth, H. (1980). *Cluster analysis algorithms for data reduction and classification of objects*. New York: Wiley.

Steinley, D. (2004). Properties of the Hubert-Arabie adjusted Rand index. *Psychological Methods*, *9*, 386–396.

Steinley, D. (2006). K-means clustering: A half-century synthesis. *British Journal of Mathematical and Statistical Psychology*, *59*, 1–34.

Symons, M. J. (1981). Clustering criteria and multivariate normal mixtures. *Biometrics*, *37*, 35–43.

Trybus, T. W., & Hopkins, L. D. (1980). Humans vs. computer algorithms for the plant layout problem. *Management Science*, *26*, 570–574.

van Oeffelen, M. P., & Vos, P. G. (1982). Configurational effects on the enumeration of dots: Counting by groups. *Memory & Cognition*, *10*, 396–404.

van Os, B. J., & Meulman, J. J. (2004). Improving dynamic programming strategies for partitioning. *Journal of Classification*, *21*, 207–230.

van Rooij, I., Stege, U., & Schactman, A. (2003). Convex hull and tour crossings in the Euclidean traveling salesperson problem: Implications for human performance studies. *Memory & Cognition*, *31*, 215–220.

Vickers, D., Butavicius, M. A., Lee, M. D., & Medvedev, A. (2001). Human performance on visually presented travelling salesman problems. *Psychological Research*, *65*, 34–45.

Vickers, D., Lee, M. D., Dry, M., & Hughes, P. (2003). The roles of the convex hull and number of intersections upon performance on visually presented traveling salesperson problems. *Memory & Cognition*, *65*, 34–45.

Vickers, D., Mayo, T., Heitmann, M., Lee, M. D., & Hughes, P. (2004). Intelligence and individual differences in performance on three types of visually presented optimisation problems. *Personality and Individual Differences*, *36*, 1059–1071.

Windham, M. P. (1987). Parameter modification for clustering criteria. *Journal of Classification*, *4*, 191–214.

Paper submitted on March 9, 2006.
The final version accepted on February 21, 2007.