
**Separation Science and Technology**
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713708471

# A New Method for Evaluating the Effectiveness of Separation Processes

W. D. Corry[a]; G. V. F. Seaman[a]; D. A. Szafron[b]
[a] Department of Neurology L104, Oregon Health Sciences University, Portland, Oregon [b] Department of Mathematical Sciences, Portland State University, Portland, Oregon

**To cite this Article** Corry, W. D. , Seaman, G. V. F. and Szafron, D. A.(1982) 'A New Method for Evaluating the Effectiveness of Separation Processes', Separation Science and Technology, 17: 13, 1469 — 1483
**To link to this Article: DOI:** 10.1080/01496398209438467
**URL:** http://dx.doi.org/10.1080/01496398209438467

SEPARATION SCIENCE AND TECHNOLOGY, 17(13 & 14), pp. 1469–1483, 1982-83

# A New Method for Evaluating the Effectiveness of Separation Processes

W. D. CORRY and G. V. F. SEAMAN

DEPARTMENT OF NEUROLOGY L104
OREGON HEALTH SCIENCES UNIVERSITY
PORTLAND, OREGON 97201

D. A. SZAFRON

DEPARTMENT OF MATHEMATICAL SCIENCES
PORTLAND STATE UNIVERSITY
PORTLAND, OREGON 97207



**Abstract**

A new method for evaluating the ability of a sorting process to attain a specific set of goals is presented. This ability is characterized by an inefficiency number, $I$, which ranges between zero and unity. Processes which fractionate objects without committing sorting errors are characterized by an $I$ of zero. Separation processes which merely subdivide a collection of objects without regard to their type are characterized by an $I$ of unity. The inefficiency number of a sorting process is calculated from observations of the number and type of errors it makes while sorting a representative collection of objects. The cost of these sorting errors is then assessed by determining how much these errors detract from the goals of the sorting procedure. This cost is then divided by the cost of randomly sorting the same collection of objects to yield the inefficiency number of the process being evaluated.

Separation processes are techniques which subdivide a mixture of objects into two or more physically separate sets. The ability of a separation technique to impart a desired order (usually spatial) to the mixture of objects sorted is the primary criterion by which the performance of the technique is measured. Unfortunately, current methods for evaluating the performance of separation techniques are rather limited. The bulk of these methods (*1–4*) have been developed to ascertain the capacity of chromatographic processes to separate a two-component mixture. Attempts to evaluate the ability of separation processes to sort multicomponent systems (*5,6*) have used entropy as a measure of performance. However, these methods cannot be easily used to evaluate the performance of a sorting process in terms of the specific goals of the procedure when more than two components are being ordered. The intent of this article is to describe two methods for evaluating the ability of a separation process to achieve a specific set of goals.

The rationale behind these techniques began with the assumption that the fewer sorting errors made while fractionating a collection of objects, the better a sorting process will be. This assumption was modified to account for the fact that not all sorting errors are equally detrimental to the goals of a sorting procedure. Thus the performance of a separation process can be measured in terms of the cost of the sorting errors made by the process while sorting a given collection of objects. The cost of a given type of sorting error refers to and is directly proportional to the extent to which it detracts from the goals of the sorting procedure. The total cost, and therefore performance of a separation procedure, is determined by the number and severity of the sorting errors made while fractionating a collection of objects. This concept forms the basis of the following methods for evaluating the performance of a separation process.



## METHOD 1

The first method for evaluating the performance of a sorting process is a general technique of almost universal application. Description of this method is facilitated by the use of the following conventions. A sorting process is regarded as a procedure for dividing a collection of $n$ objects of $k$ different types into $q$ fractions. Objects of type $i$ will have been correctly sorted only when placed into fraction $i$. A perfect sorting process is one which makes no sorting errors. A random sorting process is one which sorts a collection of objects in such a manner that the ratio of the different types of objects in each fraction is the same as in the unsorted parent sample. Equivalently, a collection of objects is randomly sorted when it is merely divided without regard for which object goes into which fraction.

The first step in evaluating a sorting procedure is to select a representative collection of objects for the procedure to sort. A representative sample is necessary in that the performance of a sorting process is often dependent upon the composition of the sample it sorts. The sample should also be well mixed to ensure that it is not ordered in a way which could affect the performance of the separation process. After the sample has been mixed, it is sorted by the process being evaluated and the composition of each fraction created is determined. The composition of any fraction, $j$, can be described in terms of a state vector, $\mathbf{S}_i$:

$$\mathbf{S}_i = S_{i1}\mathbf{e}_1 + S_{i2}\mathbf{e}_2 + S_{i3}\mathbf{e}_3 + \cdots S_{ik}\mathbf{e}_k \tag{1}$$

where $S_{i1}$, $S_{i2}$, and $S_{ik}$ are scalar quantities referring to the number of objects of type 1, 2, and $k$, respectively, which are present in fraction $i$, and $\mathbf{e}_1$, $\mathbf{e}_2$, and $\mathbf{e}_k$ are mutually orthogonal unit vectors which denote the presence of objects of type 1, 2, and $k$, respectively. After a sorting process has fractionated a given collection of objects and the composition of each fraction is determined, the distribution of objects created by the sorting process can be represented by a state matrix, S:

$$\mathbf{S} = \begin{pmatrix} S_{11} & S_{12} & S_{13} & \cdots & S_{1k} \\ S_{21} & S_{22} & S_{23} & \cdots & S_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ S_{q1} & S_{q2} & S_{q3} & \cdots & S_{qk} \end{pmatrix} \tag{2}$$

where the $i$th row represents the state vector describing the composition of fraction $i$.

After the results of the sorting process have been described in terms of a state matrix, it is necessary to assign a penalty to each sorting error made and determine the sum of these penalties. The magnitude of the penalty which is assigned to a given type of sorting error is proportional to the extent to which the sorting error compromises the goals of the sorting process. Penalties associated with the different types of errors incurred during the formation of a particular fraction can be denoted by a $k$ dimensional penalty vector, E:

$$\mathrm{E}_i = E_{i1}\mathbf{e}_1 + E_{i2}\mathbf{e}_2 + \cdots E_{ik}\mathbf{e}_k \tag{3}$$

where $\mathrm{E}_i$ is the penalty vector for fraction $i$, and $E_{i1}$, $E_{i2}$, and $E_{ik}$ are scalar quantities representing the penalties assigned to objects of type 1, 2, and $k$, respectively, placed in fraction $i$. Penalties can be assigned to the sorting errors made while creating a particular fraction and summed by taking the dot product of the state vector and the penalty vector for that fraction. This dot product represents the sum of all of the penalties due to the sorting errors made while creating that fraction. The total penalty or cost, $C$, associated with a given sorting procedure is therefore equal to the sum of the dot products for every fraction created by the process. This cost can be represented by the following equation:

$$C = \begin{array}{l} (S_{11}\mathbf{e}_1 + S_{12}\mathbf{e}_2 + \cdots S_{1k}\mathbf{e}_k) \cdot (E_{11}\mathbf{e}_1 + E_{12}\mathbf{e}_2 + \cdots E_{1k}\mathbf{e}_k) + \\ (S_{21}\mathbf{e}_1 + S_{22}\mathbf{e}_2 + \cdots S_{2k}\mathbf{e}_k) \cdot (E_{21}\mathbf{e}_1 + E_{22}\mathbf{e}_2 + \cdots E_{2k}\mathbf{e}_k) + \\ \vdots \\ (S_{q1}\mathbf{e}_1 + S_{q2}\mathbf{e}_2 + \cdots S_{qk}\mathbf{e}_k) \cdot (E_{q1}\mathbf{e}_1 + E_{q2}\mathbf{e}_2 + \cdots E_{qk}\mathbf{e}_k) \end{array} \quad (4)$$

This measure of cost can be used as an index of the ability of a separation process to achieve sorting goals, as cost is inversely proportional to performance. However, the cost of sorting a given collection of objects will vary with the size and composition of the objects sorted. These difficulties can be eliminated by normalizing the cost measured for a separation process.

The cost of a separation process, $C$, can be normalized by dividing it by the cost, $C_R$, that would be incurred had the sorting process fractionated the objects by placing the same number of objects in each fraction as before but without regard to object type. Under these circumstances the ratio of the different types of objects in each fraction is the same as in the unfractionated sample. The ratio of the measured cost of a sorting process to the cost of sorting the same objects randomly, $C/C_R$, yields a dimensionless number which will be termed an inefficiency number, $I$. A more detailed example of this method of calculating the inefficiency number of a sorting process is given in Appendix A.

The majority of separation processes are characterized by inefficiency numbers which lie between zero and unity. Separation processes which make no sorting errors warrant an inefficiency number of zero. Separation processes which sort objects randomly are characterized by an $I$ of unity. There are, however, two instances in which $I$ can assume a value greater than one or less than zero. When the cost of the sorting errors made by a separation process exceeds the cost incurred during a random sorting of the same objects, the inefficiency number of the process exceeds unity. Processes characterized by inefficiency numbers greater than unity should be avoided, as a random sorting of a collection of objects will yield results closer to those desired. Inefficiency numbers can be less than zero when negative penalties can be assigned. However, for now it will be assumed that all sorting errors are of a detrimental nature and will be assigned only positive penalties.

## METHOD 2

The second method for evaluating the performance of a sorting process is designed for situations in which it is not possible to readily determine whether sorting errors have been made. The applicability of this procedure is limited, however, to situations in which the collection of objects to be sorted contains equal numbers of the $k$ different types of objects to be sorted and the separation process can sort the collection of objects into $k$ fractions of equal size. Although these restrictions seem severe, this evaluation process is still applicable to a large variety of useful sorting techniques. A specific example, presented in Appendix B, concerns fractionation processes which sort red cell populations into equal sized fractions of increasing density.

The first phase of this method is similar to that of the first technique. A representative sample of the objects to be sorted is well mixed and fractionated into $k$ fractions. All of the objects in each fraction are then marked with a label given only to the objects in that fraction. The labeled objects are remixed and resorted as before. When this is done the composition of each fraction is described with a state vector, $\mathbf{T}_i$:



$$\mathbf{T}_i = T_{i1}\mathbf{e}_1 + T_{i2}\mathbf{e}_2 + T_{i3}\mathbf{e}_3 + \cdots T_{ik}\mathbf{e}_k \tag{5}$$

where $\mathbf{T}_i$ is the state vector for fraction $i$, $T_{ij}$ represents the fraction of cells sorted into fraction $i$ after having been initially sorted into and labeled in fraction $j$, and $\mathbf{e}_1$, $\mathbf{e}_2$, and $\mathbf{e}_k$ are mutually orthogonal unit vectors denoting the presence of cells which had initially been sorted into fractions 1, 2, and $k$, respectively. The results of the sorting process can also be fully described by a state matrix, $\mathbf{T}$, with components $T_{ij}$, which indicates what fraction of each type of labeled objects was placed in every fraction.

Penalties for sorting errors are assigned and summed by calculating the dot product of each state vector, $\mathbf{T}_i$, and penalty vector, $\mathbf{E}_i$. For every fraction produced, penalty vectors are created by the same process used in the first technique. Namely, those sorting errors which would cause the greatest damage to the goals of the sorting process are assigned the largest penalties. When all of the dot products for the process have been summed to obtain an estimate of the cost of the sorting process, the cost of randomly sorting the objects is calculated. This is easily done as each of the coefficients for the state vectors will be $1/k$, as the collection of objects sorted contained equal numbers of $k$ different types of objects and all fractions contained an equal number of objects. When the estimated cost of the sorting process is divided by the cost of randomly sorting the objects, a dimensionless ratio, $J$, is obtained. This ratio will be called an inconsistency number and it has the same general propeties as an inefficiency number. However, because the distribution of labeled objects was used to estimate the actual distribution of object types, $J(0 \leq J \leq\ )$ underestimates the actual performance of a separation process.

Although $J$ can be used as an estimate of the performance of a separation process, it is possible to obtain a much better estimate by calculating the relationship between $I$ and $J$. The relationship between $I$ and $J$ is controlled by the way in which the actual distribution of object types, **S**, determines the distribution of labeled objects, **T**.

When a sorting process fractionates a collection of $k$ different types of objects into $k$ fractions, the probability that an object of type $j$ will be placed in fraction $i$ should depend only on the type of object and the fraction involved. Thus, if $P_{ij}$ represents the probability that an object of type $j$ will be placed in fraction $i$, the composition of fraction $i$ can be characterized by the state vector, $\mathbf{S}_i$:

$$\mathbf{S}_i = P_{i1}\mathbf{e}_1 + P_{i2}\mathbf{e}_2 + \cdots P_{ik}\mathbf{e}_k \tag{6}$$

The composition of fractions can be adequately described in terms of the fraction of each type of object in them, only because every fraction is of the same size. If all of the $P_{ij}$ values $(1 < i, j < k)$ for a given sorting process are known, the distribution of the various types of objects is given by the state matrix, **S**:

$$\mathbf{S} = \begin{pmatrix} P_{11} & P_{12} & P_{13} & \cdots & P_{1k} \\ P_{21} & P_{22} & P_{23} & \cdots & P_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ P_{k1} & P_{k2} & P_{k3} & \cdots & P_{kk} \end{pmatrix} \tag{7}$$

When neither time, the sorting process, nor the labeling of the fractions created after the first sorting alter the various $P_{ij}$, it is possible to prove (Appendix C) that the actual distribution of object types represented by state matrix **S** uniquely determined the distribution of labeled objects represented by state matrix **T**. Namely, **S** times transpose **S** equals **T** or:

$$\begin{matrix} P_{11}P_{12} & \cdots & P_{1k} \\ P_{21}P_{22} & \cdots & P_{2k} \\ \vdots & & \vdots \\ P_{k1}P_{k2} & \cdots & P_{kk} \end{matrix} \quad \begin{matrix} P_{11}P_{21} & \cdots & P_{k1} \\ P_{12}P_{22} & \cdots & P_{k2} \\ \vdots & & \vdots \\ P_{1k}P_{2k} & \cdots & P_{kk} \end{matrix} = \begin{matrix} T_{11}T_{12} & \cdots & T_{1k} \\ T_{21}T_{22} & \cdots & T_{2k} \\ \vdots & & \vdots \\ T_{k1}T_{k2} & \cdots & T_{kk} \end{matrix} \tag{8}$$

Thus, if the manner (S) in which a separation process fractionates a collection of objects is known, it is possible to predict the distribution of labeled objects (T) that would occur when the fractions specified by state matrix **S** are labeled, remixed, and resorted. Unfortunately, the converse is not true. It is generally not possible to specify the particular **S** matrix which gave rise to an observed **T** matrix. However, it is still possible to estimate the inefficiency number for the separation process due to the fact that the infinitely large family of **S** matrices associated with a given **T** matrix produces a distribution of inefficiency numbers which can be characterized by a mean, standard deviation, and other statistical measures of dispersion. Thus, if all of the **S** matrices which could have given rise to an observed **T** matrix are equally probable, the mean of the inefficiency numbers associated with the family of **S** matrices is the probable inefficiency number for the observed **T** matrix. The task of estimating the probable inefficiency number of a separation process from an observed distribution of labeled objects is therefore one of generating all possible or at least a representative sample of the probable **S** matrices which could have given rise to it. The mean, minimum, and maximum of the inefficiency numbers associated with these **S** matrices provide an estimate of the true performance of the process being evaluated.

The actual method of estimating the inefficiency number of a process from its measured inconsistency number is somewhat more circuitous than the process described above. This is due to the complexity of the algorithms for generating **S** matrices which would yield an observed **T** matrix. An alternative to the procedure described above is to mathematically model the sorting process by randomly generating **S** matrices which have an inefficiency number of 0.01. As each **S** matrix is generated, the **T** matrix determined by it and the inconsistency number of the **T** matrix are calculated. When a large number (e.g., 4000) of **S** matrices with a given inefficiency number have been generated, the mean, minimum, and maximum of the inconsistency numbers of the associated **T** matrices are recorded. This process is repeated for **S** matrices which have inefficiency numbers of 0.02, 0.03, and for every other $I$ value between 0 and 1 at increments of 0.01. The minimum, maximum, and mean values of the inconsistency numbers of the **T** matrices associated with a family of **S** matrices are then plotted as in Fig. 1. Once this plot has been established, the probable inefficiency number of a process whose inconsistency number has just been measured can easily be determined. For example, if an inconsistency number of 0.20 had been measured for the process modeled in Fig. 1, the probable inefficiency number of the process would be 0.10. However, the inefficiency number of the process could have been as high as 0.125 or as low as 0.085.

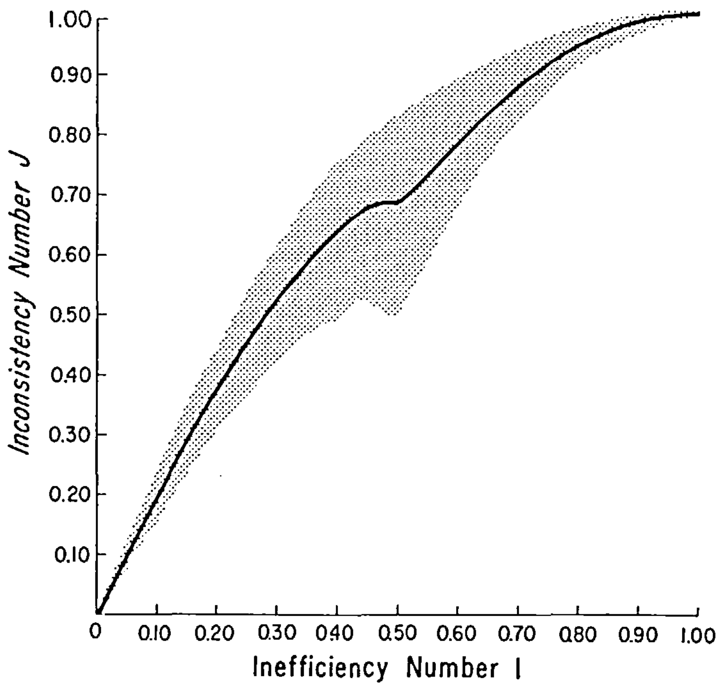


FIG. 1. Relationship between inconsistency number and inefficiency number. The relationship between $I$ and $J$ is shown here for a separation process sorting equal numbers of three kinds of objects. The penalty vectors for the three fractions are $\mathbf{E}_1 = 0\mathbf{e}_1 + 0\mathbf{e}_2 + 0\mathbf{e}_3$, $\mathbf{E}_2 = 1\mathbf{e}_1 + 0\mathbf{e}_2 + 0\mathbf{e}_3$, and $\mathbf{E}_3 = 2\mathbf{e}_1 + 0\mathbf{e}_2 + 0\mathbf{e}_3$. The solid line represents the mean value of $J$ derived from a family of **S** matrices having a common inefficiency number. The relationship between $I$ and $J$ was established at 100 incremental points between 0 and 1. At each point 4000 **S** matrices were generated to estimate the mean, minimum, and maximum values of $J$ associated with a given $I$ value. The shaded area denotes the limits of the $J$ values associated with the **S** matrices used to calculate a mean $J$ value. This particular relationship between $I$ and $J$ was generated on the condition that sortings worse than random would not occur. Thus $P_{11} > P_{12} > P_{13}$; $P_{22} > P_{12}$, $P_{32}$; $P_{33} > P_{23} > P_{13}$.

When the relationship between the inconsistency number and the inefficiency number for a given fractionation scheme has been established, the procedure for comparing the ability of two separation processes to achieve the goals of the scheme is straightforward. The inconsistency numbers for both processes are measured and with the aid of modeling studies used to determine the range of inefficiency numbers each process might have. If the range of possible inefficiency numbers associated with one separation process does not overlap that of the other process, the process with the lower probable inefficiency number is clearly superior. When the ranges of the probable inefficiency numbers for the two processes overlap, more information concerning the distribution of $I$ values associated with a given $J$ value is needed to determine whether any differences in the $J$ values of two processes are statistically significant. Information of this sort will depend on the development of algorithms for generating **S** matrices from a given **T** matrix.

In conclusion, two methods of evaluating the performance of a separation process have been presented. These methods make it possible to determine which of a variety of sorting processes is most able to meet a selected set of goals. Furthermore, these methods can improve the ability of a given sorting technique to achieve specified goals by using them to determine which modifications of the technique optimize its performance.

## APPENDIX A



As a specific example of the use of the first technique, consider the problem of evaluating the ability of a separation process to separate a mixture of three types of objects, types 1, 2, and 3, respectively, into its component parts. Assume that the sorting process fractionated a mixture of 57 objects (9 of type 1, 30 of type 2, and 18 of type 3) into 3 fractions. Fraction 1 contains 7 objects of type 1, 5 objects of type 2, and 3 objects of type 3. Fraction 2 contains 1 object of type 1, 20 objects of type 2, and 6 objects of type 3. Fraction 3 contains 1 object of type 1, 5 objects of type 2, and 9 objects of type 3. The state vectors for the fractions created by this process are:

$$\mathbf{S}_1 = 7\mathbf{e}_1 + 5\mathbf{e}_2 + 3\mathbf{e}_3 \tag{9}$$
$$\mathbf{S}_2 = 1\mathbf{e}_1 + 20\mathbf{e}_2 + 6\mathbf{e}_3 \tag{10}$$
$$\mathbf{S}_3 = 1\mathbf{e}_1 + 5\mathbf{e}_2 + 9\mathbf{e}_3 \tag{11}$$

where the coefficients in these equations correspond to the total number of each type of object in a particular fraction. The results of the sorting process can also be represented by the state matrix, **S**:

$$\mathbf{S} = \begin{pmatrix} 7 & 5 & 3 \\ 1 & 20 & 6 \\ 1 & 5 & 9 \end{pmatrix} \tag{12}$$

To assess the total cost of this sorting procedure requires a knowledge of the penalty assigned to each sorting error. Assume that the penalty assigned is proportional to the "distance" an object is from its proper fraction. For example, objects of type 3 in fraction 1 would be assigned a penalty of 2, while objects of type 1 or 3 which were found in fraction 2 would be given a penalty of 1. In that all sorting errors have a detrimental effect on the goals of this sorting process, all penalties will have a positive sign. From the aforementioned rationale for assigning penalties, the penalty vectors for this hypothetical sorting process can be written as

$$\mathbf{E}_1 = 0\mathbf{e}_1 + 1\mathbf{e}_2 + 2\mathbf{e}_3 \quad (13)$$
$$\mathbf{E}_2 = 1\mathbf{e}_1 + 0\mathbf{e}_2 + 1\mathbf{e}_3 \quad (14)$$
$$\mathbf{E}_3 = 2\mathbf{e}_1 + 1\mathbf{e}_2 + 0\mathbf{e}_3 \quad (15)$$

The total cost, $C$, incurred by this hypothetical sorting process is calculated by summing the dot products of the state and penalty vectors for every fraction. Thus,

$$C = \begin{matrix} (7\mathbf{e}_1 + 5\mathbf{e}_2 + 3\mathbf{e}_3) \cdot (0\mathbf{e}_1 + 1\mathbf{e}_2 + 2\mathbf{e}_3) + \\ (1\mathbf{e}_1 + 20\mathbf{e}_2 + 6\mathbf{e}_3) \cdot (1\mathbf{e}_1 + 0\mathbf{e}_2 + 1\mathbf{e}_3) + \\ (1\mathbf{e}_1 + 5\mathbf{e}_2 + 9\mathbf{e}_3) \cdot (2\mathbf{e}_1 + 1\mathbf{e}_2 + 0\mathbf{e}_3) \end{matrix} = 25 \quad (16)$$

The final stage in characterizing the performance of this separation process is to convert the cost of the separation process to an inefficiency number and thereby create a normalized estimate of performance. The use of a normalized estimate of cost has the advantage that its magnitude is not dependent on the size of the sample sorted, only its composition. To normalize the cost of fractionating a collection of objects, it is necessary to calculate the cost of sorting the same objects on a random basis. The state vectors describing the distribution of object types after they have been randomly sorted are determined by two requirements. Each fraction created by the random sort must have the same number of objects as the fractions created by the process being evaluated. The ratio of the different kinds of objects in these fractions must be the same as in the unsorted sample. Thus the state matrix, $\mathbf{S}_R$, for the random sorting of the objects in this example is:

$$\mathbf{S}_R = \begin{pmatrix} 2.37 & 7.89 & 4.74 \\ 4.26 & 14.21 & 8.53 \\ 2.37 & 7.89 & 4.74 \end{pmatrix} \quad (17)$$

The cost of this random sorting, $C_R$, can therefore be expressed as

$$C_R = \begin{array}{l}(2.37e + 7.89e_2 + 4.74e_3) \cdot (0e + 1e_2 + 2e_3) + \\ (4.26e_1 + 14.21e_2 + 8.53e_3) \cdot (1e_1 + 0e_2 + 1e_3) + \\ = 42.79 \\ (2.37e_1 + 7.89e_2 + 4.74e_3) \cdot (2e_1 + 1e_2 + 0e_3)\end{array} \quad (18)$$

Thus the inefficiency number for this sorting process is 0.58.

The versatility of this technique can be deduced from the observation that the technique can be used whenever the objects being sorted can be readily distinguished from one another. The nature of the sorting process is immaterial as it might be concerned with different types of cells, molecules, fibers, scrap metals, rocks, or people applying to a social club. Similarly, the penalty assigned to sorting errors are arbitrary and should reflect the goals of the sorting procedure. Thus penalties for some of these sorting procedures might be as follows: scrap metal sorting errors might be assessed in terms of dollars per pound to remove the incorrectly sorted metals, incorrectly sorted molecules might be assigned penalties reflecting the extent (ppm/mole percent/gram percent) that the errors alter the purity of the various fractions, and the penalty for incorrectly sorted people might be measured in arbitrary stress units or in terms of the loss of productivity for the other members of the club.



## APPENDIX B

As an example of the type of process the second evaluation method is designed for, consider the problem of sorting a red cell population into thirds based on the density of the cells. In that it is not currently feasible to determine the density of large numbers of individual cells, the number of sorting errors made during a fractionation procedure cannot be determined directly. Therefore, the first technique for evaluating the ability of a separation process to sort cells on the basis of their density cannot be used. The second method can be used in this instance, as there are equal numbers of each type of cell and when the cell population is fractionated into thirds, each fraction created will contain an equal number of cells.

The first step to the second method requires that the method to be tested sorts a well mixed, typical sample of red cells into three equal fractions of increasing density. The cells in each fraction are then marked with a label unique to that fraction. Tritiated glucose or $^{14}$C-glucose would provide an excellent label for this purpose. After the cells have been labeled, they are remixed and separated by the same process as before into three equal fractions of increasing density. Assume that when the distribution of labeled cell types is determined, the following T matrix is obtained:

$$\mathbf{T} = \begin{pmatrix} .4850 & .3375 & .1775 \\ .3375 & .3750 & .2875 \\ .1775 & .2875 & .5350 \end{pmatrix} \tag{19}$$

This matrix would be obtained if 48.50% of the cells initially sorted into fraction 1 were subsequently sorted into fraction 1, 33.75% were sorted into fraction 2, and 17.75% were sorted into fraction 3. Of the cells initially sorted into fraction 2, 33.75% would subsequently be sorted into fraction 1, 37.50% into fraction 2, and 28.75% into fraction 3. Similarly, of the cells initially sorted into fraction 3, 53.0% would be resorted into fraction 3, 28.75% into fraction 2, and 17.75% into fraction 1.

In creating the penalty vectors for this process, assume that the goals of the process were to place cells of type 1 in fraction 1 and that the position of the other cells was unimportant. Penalties would only be given to cells of type 1 placed in fractions 2 or 3. The magnitude of the penalties would be 2 when the cells were present in fraction 3, and 1 when present in fraction 2. Given these penalties, an estimate of the cost of the sorting process is:

$$\begin{aligned} &(0.4850\mathbf{e}_1 + 0.3375\mathbf{e}_2 + 0.1775\mathbf{e}_3) \cdot (0\mathbf{e}_1 + 0\mathbf{e}_2 + 0\mathbf{e}_3) + \\ C = &(0.3375\mathbf{e}_1 + 0.3750\mathbf{e}_2 + 0.2875\mathbf{e}_3) \cdot (1\mathbf{e}_1 + 0\mathbf{e}_2 + 0\mathbf{e}_3) + = 0.6925 \\ &(0.1775\mathbf{e}_1 + 0.2875\mathbf{e}_2 + 0.5350\mathbf{e}_3) \cdot (2\mathbf{e}_1 + 0\mathbf{e}_2 + 0\mathbf{e}_3) \end{aligned} \tag{20}$$

To calculate the inconsistency number for this purpose, it is necessary to compute the cost that would be incurred if this collection of objects were sorted randomly. Thus cost, $C_R$, is:

$$\begin{aligned} &(\tfrac{1}{3}\mathbf{e}_1 + \tfrac{1}{3}\mathbf{e}_2 + \tfrac{1}{3}\mathbf{e}_3) \cdot (0\mathbf{e}_1 + 0\mathbf{e}_2 + 0\mathbf{e}_3) + \\ C_R = &(\tfrac{1}{3}\mathbf{e}_1 + \tfrac{1}{3}\mathbf{e}_2 + \tfrac{1}{3}\mathbf{e}_3) \cdot (1\mathbf{e}_1 + 0\mathbf{e}_2 + 0\mathbf{e}_3) + = 1.00 \\ &(\tfrac{1}{3}\mathbf{e}_1 + \tfrac{1}{3}\mathbf{e}_2 + \tfrac{1}{3}\mathbf{e}_3) \cdot (2\mathbf{e}_1 + 0\mathbf{e}_2 + 0\mathbf{e}_3) \end{aligned} \tag{21}$$

Thus the inconsistency number for this separation process, $J$, is 0.6925/1.00 = 0.6925. As previously mentioned, this estimate of the performance of the separation process always makes the process seem much worse than it is.

Consider, for example, that either of the two following S matrices could have given rise to the T matrix presented.

$$\mathbf{S}_1 = \begin{pmatrix} .6000 & .2500 & .1500 \\ .3500 & .5000 & .1500 \\ .0500 & .2500 & .7000 \end{pmatrix} \tag{22}$$

$$\mathbf{S}_2 = \begin{pmatrix} .6500 & .3269 & .0231 \\ .2000 & .4808 & .3192 \\ .1500 & .1923 & .6577 \end{pmatrix} \tag{23}$$

The inefficiency numbers associated with matrices $\mathbf{S}_1$ and $\mathbf{S}_2$ are 0.45 and 0.50, respectively, and each inefficiency number is smaller than the estimate of performance provided by the inconsistency number measured for the process. To obtain a better estimate of the probable inefficiency number of the separation process which provided the observed T matrix, it is necessary to generate a mathematical model of the separation process. For the present example this model must be able to generate S matrices similar to $\mathbf{S}_1$ and $\mathbf{S}_2$, which have a predetermined (e.g., 0.01) inefficiency number. In generating these matrices, it is useful to reject matrices which correspond to physically unlikely situations. An example of a physically unlikely situation in the present case would be when there were more of the least dense cells in the heavy fraction than the light fraction and vice versa. Given this restriction, the model is allowed to generate a large number of S matrices having one inefficiency number. For each S matrix generated, the inconsistency number of the T matrix that would have been derived from it is calculated and stored. When a sufficient number of these S matrices have been generated, the mean, minimum, and maximum of the inconsistency numbers for the associated T matrices are determined and stored. This process is then repeated for S matrices having a different inconsistency number at increments of 0.01 for every inconsistency number between zero and unity. Finally, the relationship between $J$ and the probable value of $I$ associated with it can be determined for a given process by plotting the mean, minimum, and maximum values of $J$ which are associated with S matrices of a given $I$ value. This has been done for the present example in Fig. 1. Thus the probable inconsistency number for the process given in this example ($J = 0.6925$) is 0.49. However, the true value of the inconsistency number could range between 0.35 and 0.61.

## APPENDIX C

The purpose of this appendix is to prove that the relationship between state matrix **T** and state matrix **S** can be expressed as **T** = **S** times transpose **S**. This proof begins with the following assumptions.

1) The probability that an object of type $q$ will be sorted into fraction $j$ $(1 \leq j,\ q \leq n)$ by a sorting process can be represented by the quantity $P_{jq}$.
2) $P_{jq}$ is only dependent upon the type of object and the fraction involved. Neither the size of the collection of objects sorted nor its composition will alter $P_{jq}$.
3) Neither time, the sorting process, nor the label affixed to the objects in the various fractions affect $P_{jq}$.

An immediate consequence of the first assumption is that the state matrix, **S**, for a given sorting process can be written as

$$\mathbf{S} = \begin{pmatrix} P_{11} & P_{12} & P_{13} & \cdots & P_{1n} \\ P_{21} & P_{22} & P_{23} & \cdots & P_{2n} \\ \vdots & \vdots & \vdots & & \vdots \\ P_{n1} & P_{n2} & P_{n3} & \cdots & P_{nn} \end{pmatrix} \tag{24}$$

where an element, $P_{ik}$, of this matrix represents the fraction of cells of type $k$ which have been sorted into fraction $i$. A consequence of the first and third assumptions is that the probability of an object of type $q$ being initially sorted into fraction $j$ and subsequently into fraction $i$ after it has been labeled, remixed, and resorted is $Q_{ij}$, where

$$Q_{ij} = P_{iq}P_{jq} \tag{25}$$

$Q_{ij}$ can also be viewed as the fraction of objects of type $q$ which were initially sorted into fraction $j$ and subsequently into fraction $i$. Consider the sum of the fractions for every object type which was initially sorted into fraction $j$ and subsequently into fraction $i$:

$$\sum_{q=1}^{n} P_{iq}P_{jq} \tag{26}$$

Since all fractions have equal numbers of objects in them, this sum represents the fraction of objects in fraction $i$ which had previously been sorted into fraction $j$. Given this relationship, it is possible to express the distribution of labeled objects as represented by state matrix, **T**, as

$$\mathbf{T} = \begin{pmatrix} \sum_{q=1}^{n} P_{1q}P_{1q} & \sum_{q=1}^{n} P_{1q}P_{2q} & \cdots & \sum_{q=1}^{n} P_{1q}P_{nq} \\ \sum_{q=1}^{n} P_{2q}P_{1q} & \sum_{q=1}^{n} P_{2q}P_{2q} & \cdots & \sum_{q=1}^{n} P_{2q}P_{nq} \\ \vdots & \vdots & & \vdots \\ \sum_{q=1}^{n} P_{nq}P_{1q} & \sum_{q=1}^{n} P_{nq}P_{2q} & \cdots & \sum_{q=1}^{n} P_{nq}P_{nq} \end{pmatrix} \quad (27)$$

This expression for **T** is identical to the expression obtained if the matrix expressed in Relationship (24) is multiplied by its transpose, thereby proving $\mathbf{T} = \mathbf{S}\mathbf{S}^T$. It should be noted that this relationship holds only for those specific separation processes which the second evaluation process can be used on and when the elements of the state matrices represent the fraction of a particular type of object present in a given fraction.



## Acknowledgments

The authors would like to thank Drs B. Crain, E. Enneking, and L. Swanson for comments and suggestions. This work was partially supported by USPHS Grant HL 18284 for the National Heart, Lung and Blood Institute.

### REFERENCES

1. J. C. Giddings, "Optimum Conditions for Separation in Gas Chromatography, *Anal. Chem.*, *32*, 1707–1711 (1960).
2. E. Glueckauf, "Theory of Chromatography. IX. Plate Concept in Column Separations," *Trans. Faraday Soc.*, *51*, 34–44 (1955).
3. A. B. Christophe, "Valley to Peak Ratio as a Measure for the Separation of Two Chromatographic Peaks," *Chromatography*, *4*, 455–458 (1971).
4. P. R. Rony, "Extent of Separation: A Universal Separation Index," *Sep. Sci.*, *3*, 239–248 (1968).
5. G. H. Stewart, "On the Measurement and Evaluation of Separation," *Sep. Sci. Technol.*, *13*, 201–213 (1978).
6. K. de Clerk and C. E. Cloete, "Entropy as a General Separation Criterion," *Sep. Sci.*, *6*, 627–635 (1971).

*Received by editor March 5, 1982*
*Revised June 29, 1982*