PROOFS OF THEOREMS IN “A DECISION THEORETIC APPROACH TO DATA MINING” BY ELOVICI AND BRAHA, IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS. VOL. 33 (1). 2003.

### *A. Proof of Theorem 2*

We apply Blackwell’s Theorem (see Theorem 1) in a constructive fashion, namely we build stochastic matrices $M_1$ and $M_2$ such that $R \times M_1 = P$ and $R \times M_2 = Q$; thus, constituting a sufficient condition for proving the theorem.

$$P = \begin{bmatrix} p_{1,1} & \cdots & p_{1,n_y^P} \\ \vdots & & \vdots \\ p_{n_s,1} & \cdots & p_{n_s,n_y^P} \end{bmatrix} \qquad Q = \begin{bmatrix} q_{1,1} & \cdots & q_{1,n_y^Q} \\ \vdots & & \vdots \\ q_{n_s,1} & \cdots & q_{n_s,n_y^Q} \end{bmatrix}$$

$$R = Q \times P = \begin{bmatrix} q_{1,1} \cdot p_{1,1} & \cdots & q_{1,1} \cdot p_{1,n_y^P} & \cdots & q_{1,n_y^Q} \cdot p_{1,n_y^P} \\ \vdots & & & & \vdots \\ q_{n_s,1} \cdot p_{n_s,1} & \cdots & q_{n_s,1} \cdot p_{n_s,n_y^P} & \cdots & q_{n_s,n_y^Q} \cdot p_{n_s,n_y^P} \end{bmatrix}$$

Let $M_1$ and $M_2$ be two stochastic (Markov) matrices each having $n_y^P \times n_y^Q$ rows and $n_y^P$ or $n_y^Q$ columns, receptively, constructed as follows:

$$M_1 = \left[\begin{array}{ccccccc} 1 & 0 & . & . & . & . & 0 \\ 1 & 0 & . & . & . & . & 0 \\ \vdots & & & & & & \vdots \\ 1 & 0 & . & . & . & . & 0 \\ & & & & & & \\ 0 & 1 & 0 & . & . & . & 0 \\ 0 & 1 & 0 & . & . & . & 0 \\ \vdots & & & & & & \vdots \\ 0 & 1 & 0 & & & & 0 \\ \vdots & & & & & & \vdots \\ 0 & . & . & . & . & 0 & 1 \\ 0 & . & . & . & . & 0 & 1 \\ \vdots & & & & & & \vdots \\ 0 & & & & & 0 & 1 \end{array}\right] \begin{array}{l} \\ \\ \mathit{1st}\ n_y^P\ \mathit{rows} \\ \\ \\ \\ \\ \mathit{2nd}\ n_y^P\ \mathit{rows} \\ \\ \\ \\ \\ \mathit{last}\ n_y^P\ \mathit{rows} \\ \\ \end{array}$$

$$M_2 = \left[\begin{array}{ccccccc}
1 & 0 & . & . & . & . & 0 \\
0 & 1 & 0 & . & . & . & 0 \\
\vdots & & & & & & \vdots \\
0 & . & . & . & . & 0 & 1 \\
 & & & & & & \\
1 & 0 & . & . & . & . & 0 \\
0 & 1 & 0 & . & . & . & 0 \\
\vdots & & & & & & \vdots \\
0 & . & . & & & & 1 \\
\vdots & & & & & & \vdots \\
1 & 0 & . & . & . & . & 0 \\
0 & 1 & 0 & . & . & . & 0 \\
\vdots & & & & & & \vdots \\
0 & & & & & 0 & 1
\end{array}\right]
\begin{array}{l}
 \\
 \\
1st \quad n_y^Q \quad rows \\
 \\
 \\
 \\
 \\
2nd \quad n_y^Q \quad rows \\
 \\
 \\
 \\
 \\
last \quad n_y^Q \quad rows \\
 
\end{array}$$

It is easy to see that $R \times M_1 = P$ and $R \times M_2 = Q$. Hence, $R$ is generally more informative than both $Q$ and $P$.

*B. Proof of Theorem 3*

In the proof, it is assumed that investment costs $C_Q$ and $C_R$ are greater than zero. The confusion matrices $Q(C_Q)$ and $R(C_R)$ have the following general form:

$$Q(C_Q) = \begin{bmatrix}
f(C_Q) & \frac{1-f(C_Q)}{n-1} & . \quad . & \frac{1-f(C_Q)}{n-1} \\
\frac{1-f(C_Q)}{n-1} & f(C_Q) & & . \\
. & & . & . \\
. & & . & . \\
\frac{1-f(C_Q)}{n-1} & . & . \quad . & f(C_Q)
\end{bmatrix}$$

$$R(C_R) = \begin{bmatrix}
f(C_R) & \frac{1-f(C_R)}{n-1} & . \quad . & \frac{1-f(C_R)}{n-1} \\
\frac{1-f(C_R)}{n-1} & f(C_R) & & . \\
. & & . & . \\
. & & . & . \\
\frac{1-f(C_R)}{n-1} & . & . \quad . & f(C_R)
\end{bmatrix}$$

According to Blackwell's theorem [10], $Q(C_Q)$ is more effective than $R(C_R)$ if and only if there exists a stochastic matrix $M_Q$ with appropriate dimensions such that $Q(C_Q) \cdot M_Q = R(C_R)$. It is sufficient to show that $M_Q = Q(C_Q)^{-1} \cdot R(C_R)$ is a stochastic matrix while $M_R = R(C_R)^{-1} \cdot Q(C_Q)$ is not. Since $Q(C_Q)$ and $R(C_R)$ have full rank, $M_Q$ and $M_R$ can be derived.

It can be verified that $M_Q$ and $M_R$ have the following general form ($n = n_y = n_s$):

$$M_Q = \begin{bmatrix} \frac{-1+f(C_Q)+(n-1)\cdot f(C_R)}{-1+n\cdot f(C_Q)} & \frac{f(C_Q)-f(C_R)}{-1+n\cdot f(C_Q)} & \cdots & \frac{f(C_Q)-f(C_R)}{-1+n\cdot f(C_Q)} \\ \frac{f(C_Q)-f(C_R)}{-1+n\cdot f(C_Q)} & \frac{-1+f(C_Q)+(n-1)\cdot f(C_R)}{-1+n\cdot f(C_Q)} & & \vdots \\ \vdots & & \ddots & \vdots \\ \frac{f(C_Q)-f(C_R)}{-1+n\cdot f(C_Q)} & \cdots & \cdots & \frac{-1+f(C_Q)+(n-1)\cdot f(C_R)}{-1+n\cdot f(C_Q)} \end{bmatrix}$$

$$M_R = \begin{bmatrix} \frac{-1+f(C_R)+(n-1)\cdot f(C_Q)}{-1+n\cdot f(C_R)} & \frac{f(C_R)-f(C_Q)}{-1+n\cdot f(C_R)} & \cdots & \frac{f(C_R)-f(C_Q)}{-1+n\cdot f(C_R)} \\ \frac{f(C_R)-f(C_Q)}{-1+n\cdot f(C_R)} & \frac{-1+f(C_R)+(n-1)\cdot f(C_Q)}{-1+n\cdot f(C_R)} & & \vdots \\ \vdots & & \ddots & \vdots \\ \frac{f(C_R)-f(C_Q)}{-1+n\cdot f(C_R)} & \cdots & \cdots & \frac{-1+f(C_R)+(n-1)\cdot f(C_Q)}{-1+n\cdot f(C_R)} \end{bmatrix}$$

It is proved that $M_Q$ is a stochastic matrix by showing that all elements of $M_Q$ have values in the interval [0,1] and that the sum of each row equals one. First, it is shown that the sum of each row is equal to 1:

$$\forall i \quad \sum_{j=1}^{n} M_{Q_{ij}} = \frac{-1+f(C_Q)+(n-1)\cdot f(C_R)}{-1+n\cdot f(C_Q)} + (n-1)\frac{f(C_Q)-f(C_R)}{-1+n\cdot f(C_Q)} = 1$$

Next, it is shown that all the elements of $M_Q$ are positive:

$$\begin{aligned}
&C_Q > C_R \quad \rightarrow \quad f(C_Q) > f(C_R) \\
&\Downarrow \\
&-1 + f(C_Q) + (n-1)\cdot f(C_R) < -1 + n\cdot f(C_Q) \\
&\Downarrow \\
&\forall i \qquad M_{Q_{ii}} = \frac{-1 + f(C_Q) + (n-1)\cdot f(C_R)}{-1 + n\cdot f(C_Q)} \in (0,1) \\
&\forall i \neq j \qquad M_{Q_{ij}} = \frac{f(C_Q) - f(C_R)}{-1 + n\cdot f(C_Q)} \in (0,1)
\end{aligned}$$

Thus, $M_Q$ is a stochastic matrix.

Now, it is shown that $M_R$ is not a stochastic matrix. It can be verified that the sum of elements in each row of $M_R$ is 1, and that all the diagonal elements of $M_R$ are positive. We show that the off-diagonal elements of $M_R$ are negative:

$$\begin{aligned}
&C_R < C_Q \quad \rightarrow \quad f(C_R) < f(C_Q) \\
&\Downarrow \\
&-1 + f(C_R) + (n-1)\cdot f(C_Q) > -1 + n\cdot f(C_R) \\
&\Downarrow \\
&\forall i \qquad M_{R_{ii}} = \frac{-1 + f(C_R) + (n-1)\cdot f(C_Q)}{-1 + n\cdot f(C_R)} > 1 \\
&\forall i \neq j \qquad M_{R_{ij}} = \frac{f(C_R) - f(C_Q)}{-1 + n\cdot f(C_R)} < 0
\end{aligned}$$

Since $\forall i \neq j \;\; M_{R_{ij}} < 0$, $M_R$ is not a stochastic matrix. Since $M_R$ is not a stochastic matrix and $M_Q$ is a stochastic matrix, $Q(C_Q)$ is proved to be more effective than $R(C_R)$ when $C_Q > C_R$.

### *C. Proof of Theorem 4*

Our result relies on the following useful lemma:

*Lemma 1*: The matrices $S_1 = P \times R$ and $S_2 = R \times P$ are *equivalent* in terms of their effectiveness.

*Proof (Sketch)*: It is obvious that $S_1$ and $S_2$ consist of the same columns arranged in a different order. However, permutations of the columns do not affect the effectiveness of the confusion matrix, since the decision-maker needs only to permute the rows of the decision matrix accordingly.

We now prove Theorem 4. By applying Lemma 1, it is sufficient to show that $P \times Q(C_Q)$ is more effective than $P \times R(C_R)$. Since $C_Q > C_R$, by Theorem 3 $Q(C_Q)$ is more effective than $R(C_R)$. Thus, according to Blackwell's theorem [10], there exists a stochastic matrix $M$ with appropriate dimensions such that $Q(C_Q) \cdot M = R(C_R)$. The Cartesian products $P \times Q(C_Q)$ and $P \times R(C_R)$ can be represented in matrix form as follows:

$$P \times Q(C_Q) = \begin{bmatrix} P_{11} & 0 & \cdots & 0 & \cdots & P_{1n} & 0 & \cdots & 0 \\ 0 & P_{21} & & \vdots & & 0 & P_{2n} & & \vdots \\ \vdots & & \ddots & 0 & & \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & P_{n1} & \cdots & 0 & \cdots & 0 & P_{nn} \end{bmatrix} \cdot \begin{bmatrix} Q(C_Q) & 0 & \cdots & 0 \\ 0 & Q(C_Q) & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & Q(C_Q) \end{bmatrix}$$

$$P \times R(C_R) = \begin{bmatrix} P_{11} & 0 & \cdots & 0 & \cdots & P_{1n} & 0 & \cdots & 0 \\ 0 & P_{21} & & \vdots & & 0 & P_{2n} & & \vdots \\ \vdots & & \ddots & 0 & & \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & P_{n1} & \cdots & 0 & \cdots & 0 & P_{nn} \end{bmatrix} \cdot \begin{bmatrix} R(C_R) & 0 & \cdots & 0 \\ 0 & R(C_R) & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & R(C_R) \end{bmatrix} =$$

$$= \begin{bmatrix} P_{11} & 0 & \cdots & 0 & \cdots & P_{1n} & 0 & \cdots & 0 \\ 0 & P_{21} & & \vdots & & 0 & P_{2n} & & \vdots \\ \vdots & & \ddots & 0 & & \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & P_{n1} & \cdots & 0 & \cdots & 0 & P_{nn} \end{bmatrix} \cdot \begin{bmatrix} Q(C_Q) \cdot M & 0 & \cdots & 0 \\ 0 & Q(C_Q) \cdot M & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & Q(C_Q) \cdot M \end{bmatrix} =$$

$$= \begin{bmatrix} P_{11} & 0 & \cdots & 0 & \cdots & P_{1n} & 0 & \cdots & 0 \\ 0 & P_{21} & & \vdots & & 0 & P_{2n} & & \vdots \\ \vdots & & \ddots & 0 & & \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & P_{n1} & \cdots & 0 & \cdots & 0 & P_{nn} \end{bmatrix} \cdot$$

$$\cdot \begin{bmatrix} Q(C_Q) & 0 & \cdots & 0 \\ 0 & Q(C_Q) & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & Q(C_Q) \end{bmatrix} \cdot \begin{bmatrix} M & 0 & \cdots & 0 \\ 0 & M & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & M \end{bmatrix} = \left(P \times Q(C_Q)\right) \cdot \begin{bmatrix} M & 0 & \cdots & 0 \\ 0 & M & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & M \end{bmatrix}$$

$$\Downarrow$$

$$\left(P \times Q(C_Q)\right) \cdot M^{\times} = P \times R(C_R) \qquad \Rightarrow \qquad M^{\times} = \begin{bmatrix} M & 0 & \cdots & 0 \\ 0 & M & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & M \end{bmatrix}$$

Since $M^{\times}$ is a stochastic matrix, according to Blackwell's theorem $P \times Q(C_Q)$ is more effective than $P \times R(C_R)$. By applying Lemma 1 above, we conclude that $Q(C_Q) \times P$ is more effective than $R(C_R) \times P$.

*D. Proof of Theorem 5*

Let $Q$ (respectively $P$) be a confusion matrix of size $n_s^Q \times m_y^Q$ (respectively $n_s^P \times m_y^P$). Let $R = Q \otimes P$ be the doubly-Cartesian product of $Q$ and $P$. The doubly-Cartesian process operates over the set of object types in $S_R = \left\{ (s_i^Q, s_j^P) : i = 1,\dots,n_s^Q, j = 1,\dots,n_s^P \right\}$. The doubly-Cartesian process classifies object types to one of the classes in $Y_R = \left\{ (y_i^Q, y_j^P) : i = 1,\dots,n_y^Q, j = 1,\dots,n_y^P \right\}$. Thus, the doubly-Cartesian product has $n_s^Q \cdot n_s^P$ rows and $n_y^Q \cdot n_y^P$ columns.

The confusion matrices $R$, $Q$ (respectively $R$, $P$) operate on different object types and, thus, cannot be compared for their effectiveness (see Definition 1). In order to perform the comparison, the confusion matrices $Q$ and $P$ will be represented by equivalent confusion matrices $Q^E$, $P^E$ respectively, that operate on the same set of object types $S_R$:

$$Q^E = \begin{bmatrix} q_{1,1}^E & q_{1,2}^E & \cdots & q_{1,n_y^Q}^E \\ q_{2,1}^E & q_{2,2}^E & \cdots & q_{1,n_y^Q}^E \\ \vdots & & \ddots & \vdots \\ q_{n_s^Q \cdot n_s^P,1}^E & q_{n_s^Q \cdot n_s^P,2}^E & \cdots & q_{n_s^Q \cdot n_s^P,n_y^Q}^E \end{bmatrix} = \begin{bmatrix} q_{1,1} & q_{1,2} & \cdots & q_{1,n_y^Q} \\ q_{1,1} & q_{1,2} & \cdots & q_{1,n_y^Q} \\ \vdots & & \ddots & \vdots \\ q_{1,1} & q_{1,2} & \cdots & q_{1,n_y^Q} \\ \vdots & & & \vdots \\ q_{n_s^Q,1} & q_{n_s^Q,2} & \cdots & q_{n_s^Q,n_y^Q} \\ q_{n_s^Q,1} & q_{n_s^Q,2} & \cdots & q_{n_s^Q,n_y^Q} \\ \vdots & & \ddots & \vdots \\ q_{n_s^Q,1} & q_{n_s^Q,2} & \cdots & q_{n_s^Q,n_y^Q} \end{bmatrix}$$

$$P^E = \begin{bmatrix} p_{1,1}^E & p_{1,2}^E & \cdots & p_{1,n_y^P}^E \\ p_{2,1}^E & p_{2,2}^E & \cdots & p_{1,n_y^P}^E \\ \vdots & & \ddots & \vdots \\ p_{n_s^Q \cdot n_s^P,1}^E & p_{n_s^Q \cdot n_s^P,2}^E & \cdots & p_{n_s^Q \cdot n_s^P,n_y^P}^E \end{bmatrix} = \begin{bmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,n_y^P} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,n_y^P} \\ \vdots & & \ddots & \vdots \\ p_{n_s^P,1} & p_{1,2} & \cdots & p_{n_s^P,n_y^P} \\ \vdots & & & \vdots \\ p_{11} & p_{12} & \cdots & p_{1,n_y^P} \\ p_{21} & p_{22} & \cdots & p_{2,n_y^P} \\ \vdots & & \ddots & \vdots \\ p_{n_s^P,1} & p_{n_s^P,2} & \cdots & p_{n_s^P,n_y^P} \end{bmatrix}$$

According to Blackwell's theorem, in order to prove that the confusion matrix $R$ is more effective than $Q^E$ and $P^E$, stochastic matrices $M_Q$ and $M_P$ have to be found such that:

$P^E = R \cdot M_P$, and $Q^E = R \cdot M_Q$.

The following equalities can be verified:

$$Q^E = R \cdot M_Q = \begin{bmatrix} r_{1,1}^E & r_{1,2}^E & \cdots & r^E_{1,n_y^Q \cdot n_y^P} \\ r_{2,1}^E & r_{2,2}^E & \cdots & r^E_{1,n_y^Q \cdot n_y^P} \\ \vdots & & \ddots & \vdots \\ r^E_{n_s^Q \cdot n_s^P,1} & r^E_{n_s^Q \cdot n_s^P,2} & \cdots & r^E_{n_s^Q \cdot n_s^P,n_y^Q \cdot n_y^P} \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & 0 & 1 \\ 0 & \cdots & 0 & 1 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & 1 \end{bmatrix}$$

$$P^E = R \cdot M_p = \begin{bmatrix} r_{1,1}^E & r_{1,2}^E & \cdots & r^E_{1,n_y^Q \cdot n_y^P} \\ r_{2,1}^E & r_{2,2}^E & \cdots & r^E_{1,n_y^Q \cdot n_y^P} \\ \vdots & & \ddots & \vdots \\ r^E_{n_s^Q \cdot n_s^P,1} & r^E_{n_s^Q \cdot n_s^P,2} & \cdots & r^E_{n_s^Q \cdot n_s^P,n_y^Q \cdot n_y^P} \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ \vdots & & & \vdots \\ 1 & \cdots & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & 1 \end{bmatrix}$$

Since $M_Q$ and $M_P$ are stochastic matrices, the confusion matrix $R$ is more effective than both $Q^E$ and $P^E$.

### *E. Proof of Theorem 6*

Our result relies on the following useful lemma whose proof is similar to the proof of Lemma 1:

*Lemma 2*: The matrices $S_1 = P \otimes R$ and $S_2 = R \otimes P$ are *equivalent* in terms of their effectiveness.

By Lemma 2, $Q \otimes P$ (respectively $R \otimes P$) is equivalent to $P \otimes Q$ (respectively $P \otimes R$). Thus, it is sufficient to show that $P \otimes Q$ is more effective than $P \otimes R$. Let $Q$ and $R$ be the following confusion matrices operating on a finite set of object types $S$ and finite sets of classes $Y^Q$ and $Y^R$ respectively:

$$Q = \begin{bmatrix} Q_{1,1} & \cdot & \cdot & \cdot & Q_{1,n_y^Q} \\ \cdot & \cdot & & & \cdot \\ \cdot & & \cdot & & \cdot \\ \cdot & & & \cdot & \cdot \\ Q_{n_s,1} & \cdot & \cdot & \cdot & Q_{n_s,n_y^Q} \end{bmatrix},$$

$$R = \begin{bmatrix} R_{1,1} & \cdot & \cdot & \cdot & R_{1,n_y^R} \\ \cdot & \cdot & & & \cdot \\ \cdot & & \cdot & & \cdot \\ \cdot & & & \cdot & \cdot \\ R_{n_s,1} & \cdot & \cdot & \cdot & R_{n_s,n_y^R} \end{bmatrix}$$

Let $P$ be a confusion matrix operating on a finite set of object types $S^P$ and a finite set of classes $Y^P$, where $S \cap S^P = \varnothing$:

$$P = \begin{bmatrix} P_{1,1} & \cdot & \cdot & \cdot & P_{1,n_y^P} \\ \cdot & \cdot & & & \cdot \\ \cdot & & \cdot & & \cdot \\ \cdot & & & \cdot & \cdot \\ P_{n_s^P,1} & \cdot & \cdot & \cdot & P_{n_s^P,n_y^P} \end{bmatrix}$$

Let $Q^{\otimes}$ and $R^{\otimes}$ be:

$$Q^{\otimes} = P \otimes Q$$

$$R^{\otimes} = P \otimes R$$

According to Blackwell's theorem there exists a stochastic matrix $M$ with appropriate dimensions such that $Q \cdot M = R$. The doubly-Cartesian processes $Q^{\otimes}$ and $R^{\otimes}$ can be represented in matrix form as follows:

$$Q^{\otimes}=\left(P\otimes Q\right)=\begin{bmatrix}
P_{1,1} & 0 & \cdots & 0 & \cdots & P_{1,n_y^P} & 0 & \cdots & 0 \\
0 & P_{1,1} & & \vdots & & 0 & P_{1,n_y^P} & & \vdots \\
\vdots & & \ddots & 0 & & \vdots & & \ddots & 0 \\
0 & \cdots & 0 & P_{1,1} & \cdots & 0 & \cdots & 0 & P_{1,n_y^P} \\
\vdots & & & & \vdots & & & & \vdots \\
P_{n_s^P,1} & 0 & \cdots & 0 & \cdots & P_{n_s^P,n_y^P} & 0 & \cdots & 0 \\
0 & P_{n_s^P,1} & & \vdots & & 0 & P_{n_s^P,n_y^P} & & \vdots \\
\vdots & & \ddots & 0 & & \vdots & & \ddots & 0 \\
0 & \cdots & 0 & P_{n_s^P,1} & \cdots & 0 & \cdots & 0 & P_{n_s^P,n_y^P}
\end{bmatrix}\cdot\begin{bmatrix}
Q & 0 & \cdots & 0 \\
0 & Q & & \vdots \\
\vdots & & \ddots & 0 \\
0 & \cdots & 0 & Q
\end{bmatrix}$$

$$R^{\otimes}=\left(P\otimes R\right)=\begin{bmatrix}
P_{1,1} & 0 & \cdots & 0 & \cdots & P_{1,n_y^P} & 0 & \cdots & 0 \\
0 & P_{1,1} & & \vdots & & 0 & P_{1,n_y^P} & & \vdots \\
\vdots & & \ddots & 0 & & \vdots & & \ddots & 0 \\
0 & \cdots & 0 & P_{1,1} & \cdots & 0 & \cdots & 0 & P_{1,n_y^P} \\
\vdots & & & & \vdots & & & & \vdots \\
P_{n_s^P,1} & 0 & \cdots & 0 & \cdots & P_{n_s^P,n_y^P} & 0 & \cdots & 0 \\
0 & P_{n_s^P,1} & & \vdots & & 0 & P_{n_s^P,n_y^P} & & \vdots \\
\vdots & & \ddots & 0 & & \vdots & & \ddots & 0 \\
0 & \cdots & 0 & P_{n_s^P,1} & \cdots & 0 & \cdots & 0 & P_{n_s^P,n_y^P}
\end{bmatrix}\cdot\begin{bmatrix}
R & 0 & \cdots & 0 \\
0 & R & & \vdots \\
\vdots & & \ddots & 0 \\
0 & \cdots & 0 & R
\end{bmatrix}=$$

$$=\begin{bmatrix}
P_{1,1} & 0 & \cdots & 0 & \cdots & P_{1,n_y^P} & 0 & \cdots & 0 \\
0 & P_{1,1} & & \vdots & & 0 & P_{1,n_y^P} & & \vdots \\
\vdots & & \ddots & 0 & & \vdots & & \ddots & 0 \\
0 & \cdots & 0 & P_{1,1} & \cdots & 0 & \cdots & 0 & P_{1,n_y^P} \\
\vdots & & & & \vdots & & & & \vdots \\
P_{n_s^P,1} & 0 & \cdots & 0 & \cdots & P_{n_s^P,n_y^P} & 0 & \cdots & 0 \\
0 & P_{n_s^P,1} & & \vdots & & 0 & P_{n_s^P,n_y^P} & & \vdots \\
\vdots & & \ddots & 0 & & \vdots & & \ddots & 0 \\
0 & \cdots & 0 & P_{n_s^P,1} & \cdots & 0 & \cdots & 0 & P_{n_s^P,n_y^P}
\end{bmatrix}\cdot\begin{bmatrix}
R & 0 & \cdots & 0 \\
0 & R & & \vdots \\
\vdots & & \ddots & 0 \\
0 & \cdots & 0 & R
\end{bmatrix}\cdot$$

$$\begin{bmatrix}
M & 0 & \cdots & 0 \\
0 & M & & \vdots \\
\vdots & & \ddots & 0 \\
0 & \cdots & 0 & M
\end{bmatrix}=\left(P\otimes Q\right)\cdot M^{\otimes}=Q^{\otimes}\cdot M^{\otimes}$$

Thus, $R^{\otimes}=Q^{\otimes}\cdot M^{\otimes}$, where $M^{\otimes}=\begin{bmatrix}
M & 0 & \cdots & 0 \\
0 & M & & \vdots \\
\vdots & & \ddots & 0 \\
0 & \cdots & 0 & M
\end{bmatrix}$. Since $M^{\otimes}$ is a stochastic matrix, we conclude that $Q^{\otimes}$ is more effective than $R^{\otimes}$.