Suppose you have a matrices A and B that can be multiplied. Let's consider the rows of A as vectors a_i and the columns of B as vectors b_i:

$$A = \begin{bmatrix} — & a_0 & — \\ — & a_1 & — \\ — & \cdots & — \\ — & a_m & — \end{bmatrix}$$

$$B = \begin{bmatrix} | & | & | & | \\ b_0 & b_1 & \cdots & b_n \\ | & | & | & | \end{bmatrix}$$

The transposes of A and B can be written similarly:

$$A^T = \begin{bmatrix} | & | & | & | \\ a_0 & a_1 & \cdots & a_m \\ | & | & | & | \end{bmatrix}$$

$$B^T = \begin{bmatrix} — & b_0 & — \\ — & b_1 & — \\ — & \cdots & — \\ — & b_n & — \end{bmatrix}$$

Let's examine the product AB:

$$AB = \begin{bmatrix} a_0 \cdot b_0 & a_0 \cdot b_1 & \cdots & a_0 \cdot b_n \\ a_1 \cdot b_0 & a_1 \cdot b_1 & \cdots & a_1 \cdot b_n \\ \vdots & \vdots & \ddots & \vdots \\ a_m \cdot b_0 & a_m \cdot b_1 & \cdots & a_m \cdot b_n \end{bmatrix}$$

Generalizing, we can say element ij of AB is $a_i \cdot b_j$. Now, let's examine $B^T A^T$:

$$B^T A^T = \begin{bmatrix} b_0 \cdot a_0 & b_0 \cdot a_1 & \cdots & b_0 \cdot a_m \\ b_1 \cdot a_0 & b_1 \cdot a_1 & \cdots & b_1 \cdot a_m \\ \vdots & \vdots & \ddots & \vdots \\ b_n \cdot a_0 & b_n \cdot a_1 & \cdots & b_n \cdot a_m \end{bmatrix}$$

Suppose we transpose this:

$$(B^T A^T)^T = \begin{bmatrix} b_0 \cdot a_0 & b_1 \cdot a_0 & \cdots & b_n \cdot a_0 \\ b_0 \cdot a_1 & b_1 \cdot a_1 & \cdots & b_n \cdot a_1 \\ \vdots & \vdots & \ddots & \vdots \\ b_0 \cdot a_m & b_1 \cdot a_m & \cdots & b_n \cdot a_m \end{bmatrix}$$

Generalizing, we can say element ij of $(B^T A^T)^T$ is $b_j \cdot a_i$. The dot product is commutative, i.e., $a \cdot b = b \cdot a$. This means we can commute all the elements' operands:

$$(B^T A^T)^T = \begin{bmatrix} a_0 \cdot b_0 & a_0 \cdot b_1 & \cdots & a_0 \cdot b_n \\ a_1 \cdot b_0 & a_1 \cdot b_1 & \cdots & a_1 \cdot b_n \\ \vdots & \vdots & \ddots & \vdots \\ a_m \cdot b_0 & a_m \cdot b_1 & \cdots & a_m \cdot b_n \end{bmatrix}$$

Hey, this looks a lot like and is, in fact, AB. Therefore, we have:

$$AB = (B^T A^T)^T$$

If we take the transpose of both sides, we see that the transpose of a product is the product of commuted transposes:

$$(AB)^T = B^T A^T$$

The significance of this finding is that the world can't decide on how to store matrices. Some store row-major, some column-major. The good news is that if you have a matrix multiplication routine that expects one order but you have matrices of the other order, you can simply swap the operand order when you invoke the routine. The transposition happens automatically through the storage method.