Analysis of quicksort

Consider quicksort applied to n values. Let $S(n)$ be the average number of comparisons the algorithm uses. We know $S(0) = S(1) = 0$. The partition operation can be done with exactly $n-1$ comparisons. Assume that the partition value turns out to have rank q. Then we know its location, but we still have to recursivley quicksort the $q-1$ smallest values and the $n-q$ largest values. So, on average, the remaining steps of quicksort take $S(q-1) + S(n-q)$ comparisons.

The value of q is equally likely to be any integer from 1 to n. So, on average the remaining steps of quicksort take $\sum_{q=1}^{n} \frac{1}{n}[S(q-1) + S(n-q)]$ comparisons.

We claim that $S(n) \leq an \lg n$ for some constant a and $n \geq 1$. Proof by constructive induction.

Base case: $n = 1$: $S(1) = 0$ and $a \cdot 1 \cdot \lg 1 = 0$.

Induction step: Assume it holds for all positive integers less than n. Then

$$
\begin{aligned}
S(n) &= \sum_{q=1}^{n} \frac{1}{n}[S(q-1) + S(n-q)] + n - 1 \\
&= \frac{1}{n} \sum_{q=1}^{n} [S(q-1) + S(n-q)] + n - 1 \\
&= \frac{1}{n} \sum_{q=1}^{n} S(q-1) + \frac{1}{n} \sum_{q=1}^{n} S(n-q) + n - 1 \\
&= \frac{1}{n} \sum_{q=0}^{n-1} S(q) + \frac{1}{n} \sum_{q=0}^{n-1} S(q) + n - 1 \\
&= \frac{2}{n} \sum_{q=0}^{n-1} S(q) + n - 1 \;=\; \frac{2}{n} \sum_{q=1}^{n-1} S(q) + n - 1 \qquad \text{since } S(0) = 0 \\
&\leq \frac{2}{n} \sum_{q=1}^{n-1} aq \lg q + n - 1 \qquad \text{by IH} \\
&= \frac{2a}{n} \sum_{q=1}^{n-1} q \lg q + n - 1 \\
&\leq \frac{2a}{n} \int_{1}^{n} x \lg x dx + n - 1 \qquad \text{by integral bound} \\
&= \frac{2a}{n} [\frac{x^2 \lg x}{2} - \frac{x^2 \lg e}{4}] \,|_1^n + n - 1 \\
&= \frac{2a}{n} [(\frac{n^2 \lg n}{2} - \frac{n^2 \lg e}{4}) - (\frac{1^2 \lg 1}{2} - \frac{1^2 \lg e}{4})] + n - 1 \\
&= an \lg n - \frac{an \lg e}{2} + \frac{a \lg e}{2n} + n - 1 \\
&= an \lg n + [1 - \frac{a \lg e}{2}] n + \frac{a \lg e}{2n} - 1
\end{aligned}
$$

$$\leq \quad an \lg n \qquad \text{for the induction to hold}$$

If we set $a = 2/\lg e \approx 1.39$, we need $\frac{a \lg e}{2n} - 1 \leq 0$, which always holds (since $n \geq 1$). So,

$$S(n) \approx 1.39 n \lg n$$

Now that we are finished, we may realize that

$$S(n) \sim an \lg n = (2/\lg e) n \lg n = 2n \ln n$$

which is a more natural formula. Also, we could go back and do the Constructive Induction with $S(n) \leq an \ln n$, which would simplity the algebra.