Hedges and Maxson contend that weighting should only be performed if the data are noisy, i.e., homoplasious. I believe weighting is, in principle, appropriate for all data and offer the following thought experiment. Suppose we have 24 mitochondrial sequences, α-ω, 10 kb long, from which we are able to extract 527 cladistically informative positions. They give a tree for which there are 20 character-state changes supporting every clade except one, and bootstrapping supports each of these clades at >99%, except for that clade which involves sequences α, β, and γ. There are only two positions, call them 1 and 2, that address the issue of the unresolved αβγ clade. They are shown in figure 1, along with a transformation matrix that shows the distribution of the nucleotide substitutions in the other 525 positions. (In the table the numbers s/v are the number of transition changes/the number of transversion changes going from the nucleotide on the left of the row to the nucleotide at the top of the column.) Note from the matrix that (1) transitions outnumber transversions 4 to 1 (420 to 105) and (2) there are no homoplasies among the transversions, (Y = R), but that 20% of the transitions, (A = G) and (C = T), are parallel to another 20% of them. Characters 1 and 2 contradict each other. There are two possible resolutions, equally parsimonious under uniform weighting, for the αβγ clade shown in the figure. Are they equally likely? It is a choice between a tree requiring parallel T → C substitutions, which are very common, or a tree requiring parallel T → A substitutions, which are most uncommon. The former seems more probable. Weighting does nothing more than attempt to increase the odds that you select the correct tree, by assigning weights reflective of the properties of the very data being analyzed. If that consideration is apt, then weighting is generally apt. This is not to say we know how best to weight. This is not to say the expected difference will always be in the correct direction, but it should, on average, be an improvement. It may even make some differences significant that would not otherwise have been. The improvement may not always be worth the added effort, but computers make the added effort very small. If weighting is generally apt, then how noisy the data are is irrelevant.

Hedges and Maxson assert that the seed-tree topology influences the final tree’s length. This is untrue. The seed tree is used not for initial weighting but only as the starting point for searching for the best tree it can find. This is the first pass. The best tree found is then used to assign weights for a subsequent pass. In the first pass, weights, if not uniform, are computed on the basis solely of nucleotide frequencies and are thus independent of tree topology. The last pass is last because it obtained the same best tree as did the previous pass. This means that its length is determined by weights assigned on the basis of an optimum fit of the data to the tree. If one has two different
“best” trees because of different seed trees, it is simply because each was in a local optimum from which it could not escape. But the score for each best tree, being based on weights optimized for its own topology, can be legitimately compared. Weighted parsimony asks, Which tree, when it itself is used as the basis for assigning weights, is most parsimonious? There may be a tree better than that proposed by Marshall, but it remains true, on the basis of the rules used to assign weights, that his tree was preferred to the birds plus mammals tree. That does not prove Marshall is correct, but it does mean that the difference between the authors, as expressed in their letters, seems to reside solely in their readiness to weight characters and their transformations.