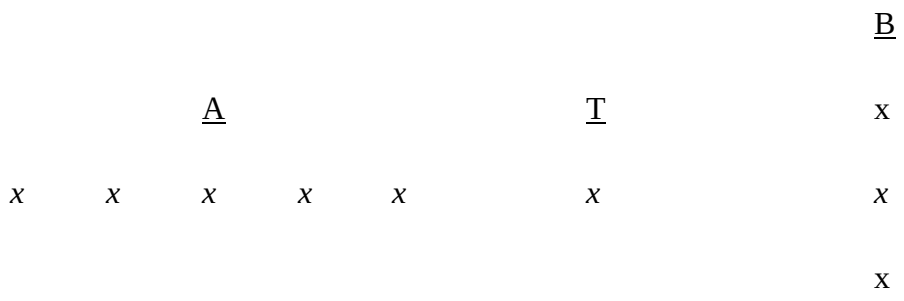
Two Concepts of Similarity

Elliott Sober

Mark Ridley, in his book *Evolution and Classification* (pp. 41-2), provides a figure like the following to illustrate the difference between two methods of phenetic clustering. The *x*'s represent taxa in a two-dimensional space (e.g., let one dimension be weight and another be height), so the distance between taxa represents overall dissimilarity. Suppose your clustering algorithm tells you that the five A taxa form a group and that the three B taxa do too. The question is where you should put T. Should T go in A or in B?



If you use the *nearest neighbor* clustering method, you'll put T with A. However, if you use the *average neighbor* clustering method, you'll put T with B. This is a problem for pheneticism. There are many ways to define overall similarity, and it is unclear how one can be correct while others are incorrect.

This problem for pheneticism is different from one that may be more familiar: Which clusters you form by using a measure of overall similarity will depend on the characteristics that comprise your data. Of course, this is true for any choice among competing hypotheses: which hypothesis you think best will depend on what your evidence is. However, the problem about pheneticism just described is supposed to involve something besides this trivial and obvious point. Pheneticism assumes that out there in nature there are groups of organisms that are united by relations of overall similarity (just as Cladistics assumes that out there in nature there are groups of monophyletic organisms). The two questions about pheneticism pertain to what this means. The point isn't merely epistemic.

Question: what does it mean for a point to be "merely epistemic?"