WILEY-BLACKWELL

Cladistics 24 (2008) 1065–1069

Cladistics

10.1111/j.1096-0031.2008.00238.x

Is REP a measure of ''objective support''?

James S. Farris[a] and Pablo A. Goloboff[b,*]

[a]*Mokekylärsystematiska laboratoriet, Naturhistoriska riksmuseet, Box 50007 SE 104-05 Stockholm, Sweden;* [b]*CONICET, Instituto Superior de Entomología, Instituto Miguel Lillo, Miguel Lillo 205, 4000 S.M. de Tucumán, Argentina*

Accepted 6 June 2008

Abstract

Grant and Kluge have recently stated that Bremer support and their own REP (''relative explanatory power''), are the only objective measures of group support. This paper discusses their claim, showing that their philosophical arguments have no basis, and that their own numerical examples actually serve to illustrate shortcomings of REP.

Introduction

According to Grant and Kluge (2008) there are only two ''objective'' measures of group support, those being Bremer support (BS; Bremer, 1988, 1994; cf. Källersjö et al., 1992; Farris, 1996) and Grant and Kluge's (2007) own Relative Explanatory Power (REP), a measure they arrived at by rescaling Bremer support. Other methods that have been used for assessing the evidence for groups, notably resampling procedures (Felsenstein, 1985; Farris et al., 1996; Goloboff et al., 2003) and relative Bremer supports (Goloboff and Farris, 2001), are ''subjective'' in Grant and Kluge's view. They advance both philosophical arguments and numerical examples, but it will soon be apparent that the former consist of little more than taking their conclusions as premises, while the latter actually reveal drawbacks of REP.

Concepts

Grant and Kluge introduce ''subjective'' as ''dependent on one's belief in the accuracy of a hypothesis,'' but that does not seem to be the actual basis of their conclusions. Consider the case of Goloboff and Farris's (2001) relative fit difference (RFD), which is defined in terms of the steps favoring (F) and contradicting (C) a group: RFD = (F − C)/F. According to Grant and Kluge, BS—which is just the difference (F − C)—is ''objective'', whereas expressing the same difference proportionally in RFD is not. Presumably the ''belief'' is introduced by the division.

To arrive at that remarkable position, Grant and Kluge devise what amounts to their own private definition of objectivity. This is based on two adequacy conditions (as they call them) for ''objective'' measures. One is:

> The second adequacy condition required for support metrics to be objective is that they quantify support as a function of explanatory power.

They identify explanatory power with parsimony (as cladists would anyway; cf. Farris, 1983, 2008), but otherwise ''a function'' is not very specific. Their other condition, however, requires that function to be proportionality in particular:

> The first is a general requirement that the metric satisfy the relation S(h | e,b) ∝ O(h | e,b), i.e. … support and optimality must vary in direct proportion to each other.

It is an unfortunate coincidence that this use of ''proportion'' occurs in a criticism of proportion RFD. To avoid misunderstanding we will refer to Grant and

*Corresponding author:
E-mail address: pablogolo@csnat.unt.edu.ar

Kluge's proportionality condition as GKP. Under GKP, then, S would be some fixed multiple of O, that is, there would be some constant k for which S = kO. To apply GKP, Grant and Kluge take BS to be a fixed multiple of optimality, then proceed to reject any measure, such as RFD, that they find is not a fixed multiple of BS.

But this means only that they chose GKP to fit their preferences; there is no other reason for it. Since logarithms are nonlinear functions, for example, it is easily seen that one measure of corroboration need not be a fixed multiple of another (Popper, 1983, p. 242):

> I prefer [formula for corroboration] D because it seems to be the simplest and most lucid of the various formulae satisfying my desiderata. But certain logarithmic formulae may do just as well—or better for certain purposes.

Nor does it seem sensible to suppose that explanatory power is a fixed multiple of optimality (parsimony), as the latter is typically measured in steps. This is apparent from Popper's (1959, p. 401) formula for explanatory power (the power of hypothesis h to explain evidence e given background knowledge b):

$$E(h, eb) = [p(e, hb) - p(e, b)]/[p(e, hb) + p(e, b)]$$

Tuffley and Steel (1997; cf. Farris, 2000) have shown that minimizing steps corresponds to maximizing the likelihood $p(e, hb)$ under their NCM model, and in that case steps S may be *monotonically* related to E, but even then the entire expression E can scarcely be a fixed multiple of S.

GKP is thus indefensible, but not all of Grant and Kluge's reasoning actually requires the strict relationship S = kO. Some of their criticisms amount simply to complaining that some other index does not have the same rank order as BS. Even this latter kind of argument, however, tacitly presumes that all indices should (or were intended to) measure the same thing as BS. Of course that assumption has nothing to do with anything that might sensibly be called "objectivity", and in any case it certainly does not hold for relative Bremer supports such as RFD, nor for the modified resampling methods we have proposed (Goloboff et al., 2003). Our intent in those approaches was to quantify the degree to which available evidence presents conflict and ambiguity in regard to the conclusion of monophyly of a given group. A group favored by 1005 characters and contradicted by 1000 has BS = 5, but its "support" is questionable at best.[1] A group favored by four characters and contradicted by none, in contrast, has lower BS but may nonetheless seem better upheld[2] by the data. Our approach to applying this idea used BS and RFD together, which is to say that we did not intend RFD as a replacement for BS. As we put it before (Goloboff et al., 2003, p. 236):

> Ideally, these two quantities should be measured separately, because they represent two aspects of the support that can vary independently.

While RFD was admittedly less than perfect, the idea that there can be more than one independent quantity to be measured, is by itself fatal to Grant and Kluge's presumption that all indices must have the same rank order. It is not surprising, then, that they never address that idea. But they evidently realize that disposing of other methods would require more than just rank order considerations, for they also try another kind of argument:

> According to Goloboff et al. (2003) and Ramírez (2005), jackknife resampling quantifies support as a relation between partitions of evidence for and against a hypothesis, whereas explanatory power is assessed in reference to the evidence analyzed as a single partition the total evidence. Jackknife resampling does not measure support as a function of explanatory power and therefore cannot be defended in terms of objective knowledge.

In fact the only partitions Goloboff et al. (2003) mentioned were splits of the tree, that is, partitions of terminals. Ramírez (2005) did mention partitioning of evidence—to point out that Grant and Kluge (2003) had been mistaken in connecting such partitioning to jackknifing! Evidently Grant and Kluge's (2008) view of "objectivity" attaches little importance to veracity.

Examples

Their other arguments being at best insubstantial, Grant and Kluge's objections to our approach rest on nothing beyond their false presumption that all our indices were intended to measure BS. That this undercuts their comments on rank order has already been pointed out, but the same error appears in their one example concerning magnitude:

> [A]s shown in Fig. 5, the RFD does not quantify support as a function of explanatory power. The RFD assigns an uncontradicted group corroborated by a single synapomorphy and an uncontradicted group corroborated by 100 synapomorphies the same support value, scoring both as maximally supported, even though the relative strength of the latter hypothesis is clearly greater.

If RFD had been intended to measure the *absolute amount* of support—as BS is supposed to do—then this behavior would indeed have been unsuitable. But RFD is instead meant to measure the *proportion* (fraction) by which characters favoring the group exceed contradic-

[1] Much the same idea was expressed by Farris et al. (1996, p. 106; see their comments on their matrix TwoX). It seems noteworthy that Kluge was a co-author of that paper.

[2] Of course in ordinary usage uphold is a synonym for support. This terminology will have to be improved, but we leave that problem for another paper.

tory ones, and the behavior that Grant and Kluge illustrate is exactly what such a proportion should do. They go on to repeat their mistake:

> The problematic behavior of the RFD is further illustrated by adding a single contradictory synapomorphy to each case. Whereas BS values indicate the relative strength of the competing hypotheses and rank clades accordingly, the RFD jumps from scoring both clades equally to ranking the two clades at opposite extremes.

Obviously, adding a conflicting character to a suite comprising just one favorable character *should* produce a large change in the *proportional* index RFD, whereas adding one character to a large suite should not.

In order to produce a legitimate criticism of our indices, Grant and Kluge would need at the least to consider the purposes for which those indices were intended, but they never address that subject. There is thus little more to say concerning their view of RFD, but it does seem worth adding that they do not appear to have devoted much thought to the behavior of their own index, REP, and it will be instructive to remedy that omission. As a first example, consider their own description of their Fig. 3:

> The upper dataset consists of five characters for five terminals. The lower dataset consists of the same characters repeated 100 times. BS values are 100 times greater for the lower dataset, but the REP support values are identical.

They regard this behavior as highly desirable, interpreting it as indicating that "REP support also allows meaningful comparison of support across datasets." But this is precisely the same behavior that they criticize when RFD shows it in their fig. 5. In Grant and Kluge's view, apparently, such behavior is objectionable only in indices proposed by others. Their fondness for double standards can be seen again from Fig. 1, which shows a situation similar to the case just discussed. As before, adding one contradictory character to the 25-character matrix on the right changes RFD from 1 to 0.96, whereas doing so to the 1-character matrix on the left changes RFD from 1 to 0. On observing the latter change in RFD, recall, Grant and Kluge describe it as extreme and problematic. They never mention that REP does the just the same. In this example REP changes from 1 to 0.92 for the large matrix and from 1 to 0 for the small one.

Again, that behavior is entirely suitable for proportional measures such as RFD; whether it is unsuitable in REP depends on the intended interpretation of REP. But considering intended interpretation seems to do REP no good, both because Grant and Kluge (2008) themselves object to such behavior and because, according to Grant and Kluge (2007, p. 486):

> REP support provides standardized values that allow support to be meaningfully compared across studies of different sets of terminals.

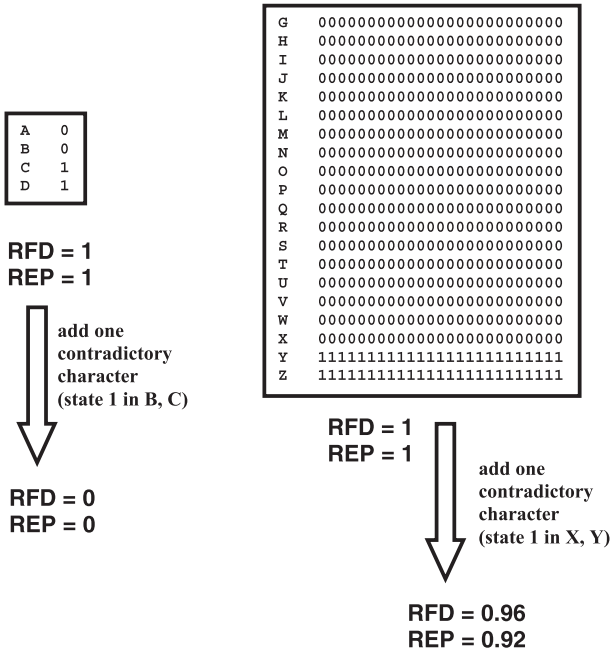

Fig. 1. Example showing that REP behaves exactly in the same way as RFD, when a single character (with state 1 in two taxa) is added to a matrix with 1 (left matrix) and 25 (right matrix) uncontradicted characters supporting a group. Grant and Kluge consider such behavior objectionable in RFD, but never mention the behavior of REP.

That wording would seem to include the present example, and in fact Grant and Kluge's way of attempting to achieve comparability of REP between data sets is itself the source of a more serious difficulty. For aligned data their standardization consists of dividing BS by the numerator of Farris's (1989) ensemble retention index R, that is by the difference $G - S$ between the length G of the bush and the length S of a most parsimonious tree. This makes REP sensitive not only to whether added characters contradict a group, but also to the number of terminals having state 0 or 1 in those characters. In the matrix of Fig. 1, REP decreases from 1.00 to 0.92 when a conflicting character with state 1 in two terminals is added, but if the conflicting character instead had state 1 in 10 terminals REP would decrease to 0.73. This effect is illustrated further in Fig. 2. Such behavior seems thoroughly unsuitable for a measure of support, since the state frequencies have no apparent logical connection with support. That criticism, it may be added, is not based only on our own view of support. It would seem to apply under Grant and Kluge's concept of support as well, since state frequencies have no such effect on BS.

State frequencies can even influence REP in the absence of character conflicts. In Fig. 3 each matrix supports a split of the terminals into two groups, and the amount of support for that split, as measured by BS, is

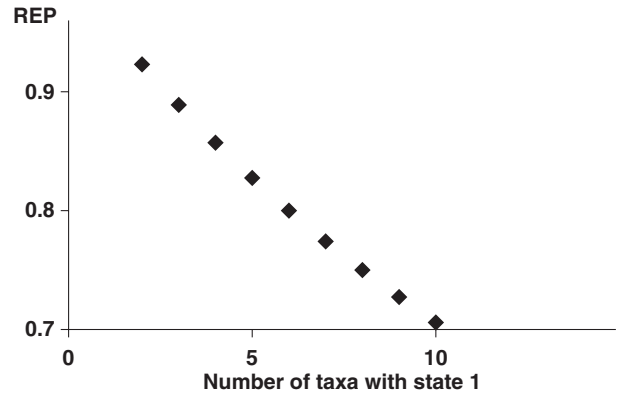

Fig. 2. Effect of the number of terminals with state 1 in a contradictory character added to the right matrix of Fig. 1. As the maximum possible numbers of steps of the contradictory character increases (maximum for an even partition, 10:10 in this case), the value of REP decreases. This leads to the conclusion that a contradictory character with apomorphic state in 10 taxa contradicts group YZ of Fig. 1 much more strongly than a contradictory character with apomorphic state in 2.

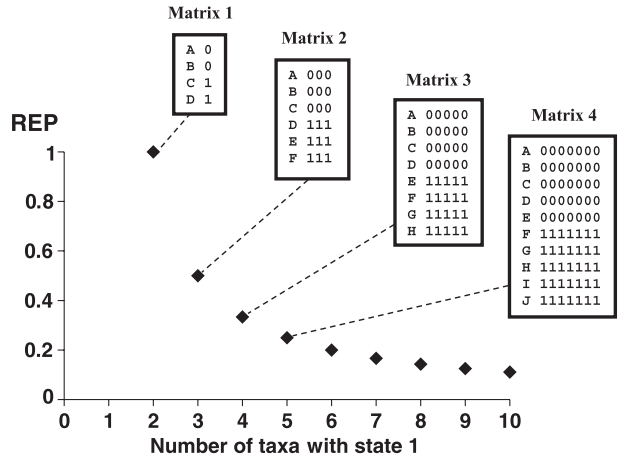

Fig. 3. Example showing that REP values of groups supported by uncontradicted characters also depends on the number of taxa with state 1. According to REP, matrix number 1 supports the monophyly of group CD much more strongly than matrix 4 supports the monophyly of group FGHIJ.

larger in the larger matrices. But REP gives just the opposite evaluation, indicating the lowest "support" in the largest matrix. Worse yet, state frequencies can interact with character conflicts. In Fig. 4, according to REP, a group set off by 15 uncontradicted characters (matrix 2) is *less* "supported" than a group favored by eight characters and contradicted by seven (matrix 1). In both these figures REP is actually negatively correlated with Bremer support!

These examples make it plain that REP is not a satisfactory measure of support, but they also bring out a more general shortcoming of Grant and Kluge's position. As Grant and Kluge present it, their approach embodies Popper's ideas:

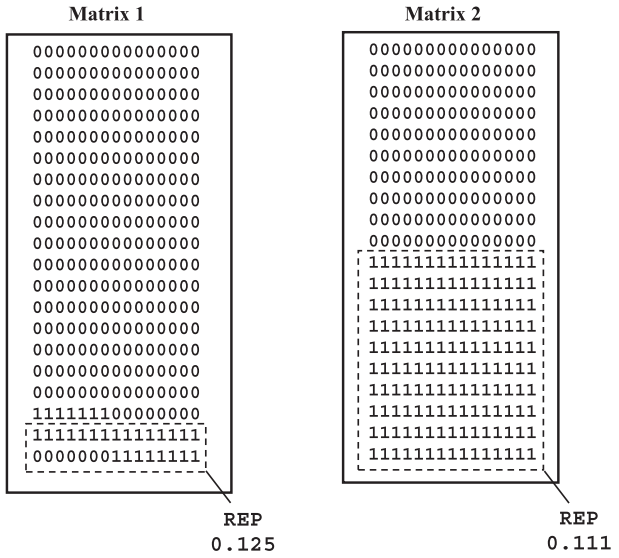

Fig. 4. A case in which REP leads to conclude that a group set off by eight characters and contradicted by 7 (matrix 1) is better "supported" than a group set off by 15 uncontradicted characters (matrix 2).

> [K]nowledge claims are objective if and only if they are open to and withstand rational criticism (Popper, 1979). Objective empirical knowledge employs deductive logic and is controlled actively by test—rational criticism involving observation and experiment.

That this is at best a pretext has already been seen, for the "objectivity" in most of their arguments is instead that defined by their private condition GKP. It may now be added that control by testing can hardly have been active in the case of REP, nor does their preference for REP seem to have involved much rational criticism. Despite their frequent appeals to "objectivity", Grant and Kluge evidently take little interest in applying Popper's criteria to their own proposals.

Acknowledgements

We thank the editor, Dennis Stevenson, for inviting a reply, and for allowing us to consider the problems posed by Grant and Kluge's arguments. We also wish to thank Salvador Arias, James Carpenter, Marcos Mirande, and Claudia Szumik for discussion. This work was supported by the ANPCyT (PICT 12605 FONCYT-BID), CONICET (PIP 02567), Consejo de Investigaciones de la Universidad Nacional de Tucumán, and Vetenskapsrådet grant VR 621-2005-4612.

References

Bremer, K., 1988. The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. Evolution 42, 795–803.

Bremer, K., 1994. Branch support and tree stability. Cladistics 10, 295–304.
Farris, J.S., 1983. The logical basis of phylogenetic analysis. In: Platnick, N.I., Funk, V.A. (Eds.), Advances in Cladistics II. Columbia University Press, New York, pp. 7–36.
Farris, J.S., 1989. The retention index and the rescaled consistency index. Cladistics 5, 417–419.
Farris, J.S., 1996. Names and origins. Cladistics 12, 263–264.
Farris, J.S., 2000. Corroboration versus ''strongest evidence''. Cladistics 16, 385–393.
Farris, J.S., 2008. Parsimony and explanatory power. Cladistics 24, 825–847.
Farris, J., Albert, V., Källersjö, M., Lipscomb, D., Kluge, A., 1996. Parsimony jackknifing outperforms neighbor joining. Cladistics 12, 99–124.
Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783–791.
Goloboff, P., Farris, J., 2001. Methods for quick consensus estimation. Cladistics 17, S26–S34.
Goloboff, P., Farris, J., Källersjö, M., Oxelmann, B., Ramírez, M., Szumik, C., 2003. Improvements to resampling measures of group support. Cladistics 19, 324–332.
Grant, T., Kluge, A.G., 2003. Data exploration in phylogenetic inference: scientific, heuristic, or neither. Cladistics 19, 379–418.
Grant, T., Kluge, A., 2007. Ratio of explanatory power (REP): a new measure of group support. Mol. Phylogenet. Evol. 44, 483–487.
Grant, T., Kluge, A., 2008. Support metrics and their adequacy. Cladistics (in press) doi: 10.1111/j.1096-0031.2008.00231.x.
Källersjö, M., Farris, J.S., Kluge, A.G., Bult, C., 1992. Skewness and permutation. Cladistics 8, 275–287.
Popper, K.R., 1959. The Logic of Scientific Discovery. Hutchinson, London.
Popper, K.R., 1979. Objective Knowledge: An Evolutionary Approach. Oxford University Press, New York.
Popper, K.R., 1983. Realism and the Aim of Science. Routledge, London.
Ramírez, M.J., 2005. Resampling measures of group support: a reply to Grant and Kluge. Cladistics 21, 83–89.
Tuffley, C., Steel, M., 1997. Links between maximum likelihood and maximum parsimony under a simple model of site substitution. Bull. Math. Biol. 59, 581–607.