Since three different individuals asked this weekend about whether the SOP formula would be a superior method of doing elim breaks and seeding as opposed to the traditional method of using speaker points with various adjustment tiebreakers, I thought I should at least explain the alternatives.

First, a quick clarification – the SOP method for creating high-low pairings is not based on the premise that SOP creates a more accurate seeding to determine the "best" and the "worst" team in the bracket. SOP is designed to equalize strength of opposition within win/loss brackets to maximize fairness by providing an equal opportunity for success. The team at the top is largely there not because they are "best" but rather because thus far they have been "unluckiest." Similarly, the team at the bottom of an SOP bracket is there primarily because they have been lucky thus far. So SOP gets balanced iteratively over the course of the tournament.

So it is not accurate to say, "we've decided that SOP is the best way to seed teams during prelims, why don't we do it that way during elims." SOP has a different rationale for prelims that doesn't entirely apply in elims. That said, it is indeed possible to argue pro and con as to whether SOP formulas that incorporate strength of opposition create a more accurate (or more fair) seed order for determining breaks and seed position in elims as well.

One more preliminary issue before we get to the data. It should be noted that the SOP model is not principally different in that it includes strength of opposition while others don't, but more in HOW it does it. There are two major ways to deal with a set of different ranking statistics. We'll use pro football and college football as the example. In the NFL, seeding is based on an ordered set of tie-breakers that are each considered in turn with lower ranked tiebreakers considered only if higher ranked tiebreakers are all tied. That's how we do seeding in debate at present. So if we include strength of opposition, it would rarely apply because several others things would have to be tied first. But in the BCS, the ranking is a single aggregated statistic that incorporates computer models, polls, etc. into a weighted number that can serve as a ranking mechanism. That's what SOP does by combining my seed with the average seed of my opponents to make a single composite statistic. Now there are a huge number of ways to create ordered lists of tiebreakers but an even larger (infinite) array of possible formulas to weight and aggregate variables. In the end, the latter class of models is more robust than the former but they can lose any sense of "reality." They are just abstractions which may or may not remain meaningful to the participants. Even with the complexity of NFL tiebreakers, most observers at least know why and how the ties get broken. BCS computer models can be very obscure and non-reproducible. SOP is a rather modest proposal since it only aggregates two statistics, both of which are derived from the traditional ranking mechanisms that we've always used (seeds are still generated by a ranked list of tiebreakers). In fact, critics of SOP wonder how we can "reject" HL seeding using speaker points (I don't) and then turn around and use points to calculate my seed and that of my opponents.

With that said, let's address the data. The question was asked – how would seeds be different if we seeded on SOP instead of our traditional speaker point tiebreakers? The short answer,

they would be different but not by much – though if you are a team that would break by one model and not by the other, the difference appears HUGE. At GSU, exactly the same list of 32 teams would have cleared, though not in the same order. At Kentucky, three teams that didn't clear would have replaced teams that did. The correlation between seeds calculated in the traditional way and with SOP proved to be .991 for the 91 teams at GSU and .988 for the 148 teams at Kentucky.

But if the seeds are actually different, meaning that different teams might clear and different debates occur among the teams that do clear, which seeding is "correct" (if either)? It would be nice to say that this is an empirical question for which I could make a statistical argument. But I can't. It is primarily a philosophical and a conceptual argument rather than an empirical one. There really isn't any demonstrably "correct" seed order. There are only the seed orders which the community defines as most fair and which correctly reward the performances that we want to reward. In my personal opinion, SOP probably deserves more consideration than being the 3rd or 4th or 5th tiebreaker. My bias might also say that strategies that aggregate statistics rather than order tem are better. Why, for instance, do we privilege HL adjusted points as the single most important statistic after WL? But that's for the community to decide. I suspect that I'm not being asked to state my beliefs but rather to make some sort of statistical demonstration.

If we were to try to make empirical arguments, there's a couple of places we might look. For instance, if seeds are to be "predictive" of the outcome of a debate (a different topic for discussion), one could look specifically at the debates that occurred in elims, comparing how the teams were seeded as opposed to what they would have been seeded with SOP. There were four elim rounds at Kentucky where SOP would have created seeds where if the debate occurred a different team would have been favored than was favored using the assigned seeds. Presumably, if SOP was more accurate, one might guess that the outcome of those four rounds would have followed the SOP seeds rather than the actual seeds. But in 3 of the 4 rounds, the team that had the actual higher seed won. Now this isn't much of an argument, since it could be assumed that since people know the actual seeds, they have a slight perceptual advantage. True enough. But given a burden of proof on change, there isn't really any evidence that the SOP seeds more accurately "predicted" who should win the round.

A second line of argument would be to look at upsets. Since upsets could be used as a disconfirmation of the predictive power of the assigned seeds (not really but so goes the argument), one might expect that the difference between the SOP seeds (or whatever more accurate seeding we used) would be less than the actual seeds so that the upset would seem less anomalous. But as noted below, this wasn't the case at either Kentucky or GSU.

At Kentucky there were 8 "upsets" out of 31 elim debates. In one debate the SOP seed would have reversed the upset by actually favoring the team that won. But, of course, as noted above it would have created 3 more putative upsets by incorrectly predicting that a higher seeded SOP team should have won when they didn't. For the rest of the rounds where there isn't a reversal of who's favored, you can only look at how "large" the upset appeared to be in seed

terms. The difference was often GREATER rather than LESS using SOP. So it would be marginally less predictive.

It is also interesting to track the success of the 9 seed through three upsets all the way to the finals. It is a not particularly unusual story where a high 6 defeats teams with more wins but fewer points. Points actually become quite predictive in elims, with probably more overall predictive power than wins (a study I haven't done formally but have observed repeatedly over the years).

Actual	SOP
27 v 6 (21)	29 v 7 (22)
26 v 7 (19)	34 v 5 (29)
23 v 10 (13)	23 v 19 (4)
17 v 16 (1)	11 v 14 (-3)
9 v 8 (1)	13 v 12 (1)
9 v 1 (8)	13 v 1 (12)
5 v 4 (1)	6 v 4 (2)
9 v 5 (4)	13 v 6 (7)

The final piece of data is that the 26th seed in actual terms would not have cleared using SOP. But they not only cleared but won. So it "could" be argued that their actual seed underestimated their strength rather than overestimated it if we assumed SOP was right. Their outcome in doubles seemingly confirmed that the "belonged." Of course, anecdotes cut both ways. If I'm the team that loses to them, I'm tempted to say, "they shouldn't have even cleared." But that becomes viciously circular.

At GSU there were also 8 "upsets" out of 31 elim debates in open. In only one of the eight debates was the seed difference less using the revised seeds as opposed to the actual seeds and that difference was only 1 (12 as opposed to 13). While not statistically significant, the actual seeds actually did a better job of prediction.

31 v 2 (29)	32 v 1 (31) (the same debate would have happened either way)
22 v 11 (11)	28 v 7 (21)
18 v 15 (3)	20 v 10 (10)
17 v 16 (1)	17 v 13 (4)
31 v 18 (13)	32 v 20 (12)
13 v 4 (9)	12 v 3 (9)
9 v 8 (1)	9 v 6 (3)
7 v 3 (4)	8 v 4 (4)

Overall, I couldn't say that this data provides a strong empirical rationale that traditional seeds are sacred. But I'd have an even more difficult time saying that SOP seeds are empirically better. In the end, a completely different kind of argument needs to be made.