

May 6, 2005

Dear Dr. Rayner,

Per your instructions, we have revised our manuscript (MS#2004-0163R) along the lines requested by reviewer 2. In the paragraphs below, we summarize these changes as well as address the comments of reviewer 1.

Reviewer 2:

Reviewer 2 had two primary concerns. The first was to temper conclusions about the generality and robustness of PSP. We have done this throughout the manuscript. We also believe that this reviewer will find that claims regarding the robustness of the volume estimation procedure, which s/he felt were too simplistic and biased in favor of PSP, more justifiable in light of new developments in this project. Since the resubmission of the manuscript on 2/17/05, we have developed a more accurate volume estimation method. Because it is a substantial improvement over the prior method, the volume estimation test in Appendix B was redone. It was not only more rigorous so as to be in line with reviewer 2's wishes, but the results nicely demonstrate good accuracy. Because we could not make this change without also updating the three application examples in the main body of the manuscript, all of the volume estimates in these simulations were also redone with the updated method. This was an enormous amount of work, and is why the revision is a few days late. The new estimation method changed the outcomes in a few places (generally shrinking volumes), so the text was updated accordingly.

Reviewer 2's second concern can be broken down into two parts, the first is minor and the second is more substantive. S/he first asks whether a simpler search method might succeed as well as PSP in finding the different patterns. We have a few responses to this question. In the toy example (Table 1 and Figure 5) we used SMC (simple monte carlo) as an alternative search method to show that random search not only fails miserably but can be so slow as to be useless. When applied to real models like TRACE and Merge, the same failures are observed. For example, in the simpler of the two experimental designs in which these models were compared, Frauenfelder et al (1990), SMC failed to find some of the patterns that PSP had no difficulty in discovering. In five replications of the search of TRACE's parameter space (weak threshold), SMC always found only 9 patterns during a 500-second test, whereas PSP found all 12 during the same time period. SMC could not find the additional 3 pattern even when the search was prolonged several hours.

We have not experimented with other search methods. The difficulty is in finding an efficient and "smart" search method that does not miss regions when applied in high-dimensional space. PSP's methodical mapping of the search space minimizes the likelihood of this problem. In this regard, mapping every region, although costly in terms of efficiency (time), ensures a much more successful search in the end.

Reviewer 2 also wanted a more thorough discussion of the issues underlying the volume analysis. We are grateful for this request because it revealed a need for us to distinguish more clearly between the counting analyses and volume analyses. Volume estimation is an informative by-product of a PSP search, from which much can be learned about a model. However, as we

now make clear in the manuscript (e.g., pp. 23, 51-52), volume estimation is specific to a parameterization of a model, whereas counting the number of patterns a model generates is not. That is, regardless of how a model is parameterized, it will always generate the same number of data patterns; it is invariant under reparameterization. PSP was purposefully developed to have this property, which is also true of its predecessors, like MDL, and makes the technique extremely powerful.

Volume estimation is not invariant under reparameterization. By changing the functional form of a model, volume estimation might change. Therefore, for a volume analysis to be meaningful, the modeler must be willing to attach meaning to the parameters (or other important aspects of the models), which usually involves associating them with theoretical/psychological constructs. Without this additional step, a region's volume has no meaning. In practice, this additional step is not an issue, as it is a precondition for being a model of cognition as opposed to just an equation. What makes modeling meaningful is that a model's parameters are linked to underlying mental processes (Estes, 2002). Without this step, the enterprise of cognitive modeling would degenerate to no more than curve fitting.

Reviewer 1:

We hope it is obvious from how we have revised the manuscript that we try to be responsive to reviewers' concerns. With regard to reviewer 1, there is such a stark difference of opinion on the theoretical value of the research that it is difficult to be similarly responsive. For reviewer 1, it seems that only a manuscript that would use PSP to resolve a substantive theoretical debate in a particular content area would make for a publishable manuscript in *Psychological Review*. This viewpoint ignores the fact that we are addressing central theoretical issues that are broader than a particular content area. They include questions like, How should we decide between competing models? How do we assess model similarity/mimicry? How powerful is a model? To what extent is a model's account of human data robust and sensible? Do structural and processing differences between models make them functionally different? These are overarching questions that all of psychology faces, and will face more and more as modeling continues to grow in the various subfields. PSP represents a substantial advance in addressing them. It was motivated from theoretical questions, like the ones above, that we ask ourselves during the course of our own modeling efforts. PSP is much more than a "sensitivity analysis," as reviewer 1 states.

In general, we feel that this reviewer's criticisms attempt to diminish the contribution and significance of our work. In his/her first point, this reviewer trivializes PSP as being a minor enhancement on past work that makes it more appropriate for *Psychological Methods* or *Behavior Research Methods, Instruments & Computers*, when in reality PSP is a significant advance in model evaluation, comparison, and testing. What is learned about model behavior in the three applications examples is clear evidence of this. PSP offers keen insight into model performance and selection. There is no other method that has the power or versatility of PSP. If one of these other publication outlets were more suitable, we would have sent it there in the first place, and not have wasted your or the reviewers' time if we truly felt the paper belonged in one of them.

In reviewer 1's second point, s/he makes demands that could not be met by any model analysis method. To expect PSP to be universal, applicable to all models, and full proof is an

impossibility. No numerical method today of any sort lives up to this standard. It does not seem to matter to reviewer 1 that PSP is applicable to “some of the most widespread and influential frameworks in cognitive modeling,” just not the one that matters most to this reviewer. That reviewer 1 faults PSP for the assumptions that must be satisfied to guarantee accurate performance (continuity, contiguity, stationarity) ignores the fact that all numerical implementations of statistical methods, including statistical models, have limiting conditions and therefore require assumptions. It is necessary to be up front and clear about these. Indeed, reviewer 2 seems to appreciate our attention to detail. Most cognitive models are designed such that they are likely to obey the two most critical assumptions, continuity and contiguity. Even if they do not, there are ways to test for violations, as we discuss at length in Appendix A and in the General Discussion. Also, the modeler is likely to know the model well enough to suspect violations and be alert to them. All this said, PSP continues to evolve to overcome limitations, of which our improved volume estimation method is the most recent example.

In the last revision, we expanded the discussion of parameter choice, but apparently did not give reviewer 1 what s/he wanted. We have again tried to be more responsive in this second revision (pp. 50- 51).

Finally, both reviewers’ specific comments were addressed by augmenting and clarifying the relevant sections of text with an answer, as other readers might have similar questions.

We thank you and the reviewers for a timely turn-around on this last revision, and hope to hear from you soon on the current one.

Sincerely

Mark Pitt, Woojae Kim, Daniel Navarro, Jay Myung