Title: Are divergence point analyses suitable for response time data?

Authors: Gomez, Breithaupt, Perea & Rouder

Reviewer: Eyal Reingold

The stated goal of the submitted manuscript was to evaluate the properties of a novel distributional analysis technique we recently introduced for deriving divergence point estimates (Reingold, Reichle, Glaholt & Sheridan, 2012; Reingold & Sheridan, 2014). In their critique the authors make extremely strong negative proclamations about our procedure. In reading the paper I was struck by the arrogant tone and by the brazen distortion and misrepresentation of our work. Importantly, while their conclusions are extremely broad and authoritative, the authors are very evasive and vague about the logical and empirical basis for these claims.

Here I will only address what I understand to be the 2 major claims in the submitted paper:

1) The divergence point estimates are biased with low experimental power –

This is not a new revelation or insight. In Reingold and Sheridan (2014), we published an extensive exploration of this very issue, proposed modification for our original procedure (Reingold et al., 2012), and demonstrated by using several simulation the robustness of these modified procedures in providing unbiased estimates despite low experimental power. The authors cited this paper but pretended that they are providing an important new contribution with their rather simple simulation the details of which are very sketchy. Furthermore, in this paper we explained the fact that the bias was built into the original procedure for a very important theoretical reason:

“It is important to note, that [Reingold et al. (2012)](http://journal.frontiersin.org/article/10.3389/fpsyg.2014.01432/full#B24) and follow-up investigations (e.g., [Glaholt and Reingold, 2012](http://journal.frontiersin.org/article/10.3389/fpsyg.2014.01432/full#B8); [Sheridan and Reingold, 2012a](http://journal.frontiersin.org/article/10.3389/fpsyg.2014.01432/full#B29),[b](http://journal.frontiersin.org/article/10.3389/fpsyg.2014.01432/full#B30);[Sheridan et al., 2013](http://journal.frontiersin.org/article/10.3389/fpsyg.2014.01432/full#B28); [Glaholt et al., 2014](http://journal.frontiersin.org/article/10.3389/fpsyg.2014.01432/full#B7); [Reingold and Glaholt, 2014](http://journal.frontiersin.org/article/10.3389/fpsyg.2014.01432/full#B23)) attempted to test the validity of the direct cognitive control hypothesis that predicted early divergence points as a function of cognitive influences. Consequently, in order to protect against making a Type I error (i.e., erroneously estimating a divergence point prior to the actual point of divergence), the DPA procedure incorporated very conservative criteria for estimating divergence points (i.e., α < 0.001 for the significance of individual bins and the requirement for five consecutive significant bins). Importantly, despite this deliberate bias against the direct cognitive control hypothesis, the above studies documented fast acting cognitive influences in strong support for that hypothesis (for a review see [Reingold et al., in press](http://journal.frontiersin.org/article/10.3389/fpsyg.2014.01432/full#B25)). However, the cost of such a conservative bias in the DPA procedure is the risk that the estimate of the divergence point would be delayed relative to the actual point of divergence. This would be especially the case under low experimental power (i.e., a small number of participants and observations). To mitigate this risk, the above investigations employed a large number of observations and participants. Nevertheless, it would be desirable to construct a version of the DPA technique that can successfully handle the lower experimental power that is typical of many eye movements and reaction time experiments.”(See full text at <http://journal.frontiersin.org/Journal/10.3389/fpsyg.2014.01432/full>)

In Reingold and Sheridan (2014), we provided MATLAB code for all versions of the procedure as well as a database with a built in divergence point for each participants. Our simulations demonstrated the bias in the original procedure (which was “rediscovered” by the authors’ simulation) and demonstrated the validity and reliability of the modified procedures. Importantly, by reanalyzing the data from our studies using the original procedure we got similar estimates of divergence points thereby establishing that with the increased experimental power our original procedure performed well despite the built in conservative bias. I know that the authors were aware of the Reingold and Sheridan (2014) paper not only because they cited it but also because in their discussions with my former student and coauthor Heather Sheridan they were told about it and about its relevance to their claims. The fact that they chose to blatantly ignore the contents of this paper can only be construed as an unprofessional and deliberate misrepresentation of our work. There are many other misrepresentations (including both omission and commission errors) in the description of our work. For example, reading the paper and looking at the simulation the uninformed reader would get the impression that our procedure was primarily developed and implemented to analyze RTs. In contrast, all prior studies with the exception of one recent study (in which our procedure was used as a secondary analysis of RT data) analyzed fixation duration data. In their paper the authors even erroneously refer to Sheridan (2013) as an RT study. There are too many such problems to discuss here but I believe that the above will suffice to illustrate my concern.

2) The procedure is logically flawed and makes unwarranted assumptions without a specific computational model of the underlying processes –

The arguments are convoluted and difficult to follow and evaluate. In addition, the simulations that are described as the basis for some of the claims are poorly specified (no detailed method section or appendix and MATLAB code, etc.). Without access to such information, I’m not able to evaluate the accuracy of the implementation of these simulations. Furthermore, visually comparing the density plots in Figures 2-3 indicates that they are very similar. This is extremely strange, because a pure Mu effect (Figure 3) should produce a shift only, but instead the density plots also show a substantial increase in skew (i.e., a Tau effect). Conversely, a pure Tau effect (Figure 2) should manifest as an increase in skew without a shift and this does not seem to be the case. Ironically, in many of our papers we analyzed the data using both the ex-Gaussian procedure as well as our divergence point analysis and we explicitly discussed and demonstrated the convergence between these procedures. Such convergence is in direct opposition to the claims in the present paper. For example, in the paper which first introduced the procedure (Reingold et al., 2012), we demonstrated a substantial delay in the estimate of the divergence point in a condition with a pure tau effect (invalid preview condition) as compared to a condition which also produced a mu effect (valid preview condition). Given that this is clearly inconsistent with the results reported by the authors one might reasonably expect that they would mention and discuss our work using this technique. However, once again they conveniently ignored aspects of our work that did not fit with their narrative. Furthermore, unlike the simulations we conducted (Reingold & Sheridan, 2014), the present simulations did not provide a realistic database which resembles actual fixation duration data. For example, the absence of simulated individual participants does not permit the computation of divergence point estimates using the individual participant version of our procedure. Most importantly, we provided well documented simulations as well as MATLAB code and database (Reingold & Sheridan, 2014) in order to allow researchers who would like to propose alternative methods to use our data and scripts to contrast their approach with ours. Despite this complete transparency, the authors ignored our simulations and our modified procedures.

On a more conceptual level. The authors seem to suggest that a prerequisite for the usefulness of a procedure like ours is an explicit set of processing assumptions preferably in the form of a computational model. They also attribute to us the assumption that a divergence point estimate corresponds to the onset of a differentiating process and that prior to the divergence point the processing across conditions is identical. This is a straw man of their own creation. Specifically, our method is entirely agnostic with regard to the nature of the processing differences between conditions either before or after the divergence point. It is only meant to provide an estimate of the earliest point at which an experimental manipulation produces a discernable difference between conditions. By their own admission the processing model they simulated is very simplistic. Ironically, one of the models given by the authors as an example of what we should aspire to is the E-Z Reader model by Reichle and colleagues. Using this model Sheridan and Reichle (2015) demonstrated simulations which produced divergence point estimates that were very similar to the ones we obtained empirically (another paper conveniently ignored by the authors). Furthermore, requiring an explicit processing model as a prerequisite for any statistical analysis approach is a very unusual criterion. In fact, one of the advantages of the bootstrapping approach which we incorporated is that it makes fewer assumptions than the traditional parametric statistical methods. Yet I assume that the authors would not want to propose that these methods require a processing model. Instead, we argued that the estimates derived from our procedure in the context of theoretical and computational models might provide very valuable empirical data for evaluating such models.

In addition, the authors seem to be implying that an extremely rapid divergence point is somehow “nonsensical”. Furthermore, they seem to suggest that the procedure is doomed to always produce an early divergence point estimate (because of stochastic dominance) except in the case of low experimental power (few observations per condition). In contrast, a rapid divergence point is in no way “nonsensical”. For example, a Mu difference will produce a shift of one distribution relative to another and is commonly regarded as a rapid effect that might impact the entire distribution. The divergence point estimates in that case would be expected to correspond to left tail of the distribution (i.e., the fastest durations). Perhaps the authors want to argue that early divergence points indicate a differentiation that is too fast to be plausible given our knowledge concerning perceptual, cognitive and neural delays. This issue was discussed in detail in our papers (e.g., [Reichle & Reingold, 2013](http://journal.frontiersin.org/article/10.3389/fnhum.2013.00361/abstract); Reingold et al., 2012). For example in Reingold et al. (2012) we demonstrated that rapid divergence point estimates occur due to parafoveal processing of target words prior to the first-fixation on these words and that when parafoveal processing is rendered ineffective divergence points occur much later. Thus, preprocessing of stimuli and/or differences in “set” might result in extremely rapid divergence point estimates and we think this is entirely consistent with theoretical models. Just like with any methodology the more similar the conditions that are contrasted the more meaningful is a finding of a difference between these conditions. If the authors want to simply argue that when distributions are largely non-overlapping the divergence point analysis might not be as informative then I don’t have much of a disagreement with them. However, many of the experimental manipulations we analyzed have much more subtle effects on fixation duration data and even in the case of manipulations producing large effects, there is often substantial variation in divergence point estimates across participants (see Reingold & Sheridan, 2014). Specifically, both our data and simulations demonstrated a wide range of divergence point estimates (using both analyses based on individual participant data as well as group based data), findings that appear to contradict the authors’ assertions concerning the inevitable stochastic dominance.

In summary, the authors seem to agree that our goal of obtaining an estimate of the divergence point is important. Despite this agreement, their critique of our approach is not accompanied by any suggestion of a better method. More seriously, this critique seems to be based on a grossly distorted caricature of our approach, vague arguments, and poorly described simulations. Consequently, I find the paper to be neither insightful nor useful. I’m sorry to say that I do not see any contribution here that would merit publication anywhere let alone in Cognitive Psychology.

Sincerely,

Eyal Reingold