UNIVERSITY OF SOUTHERN MAINE
PORTLAND • GORHAM • LEWISTON • ONLINE

Jeffrey A. Walker
Professor of Biological Sciences
Department of Biological Sciences
70 Falmouth St
Portland, ME, 04103
Phone: (207) 228-8166
E-mail: walker@maine.edu
URL: http://usm.maine.edu/bio/jeffrey-walker

August 2, 2016

Dear Editors,

I thank the reviewers and editor for their constructive comments on my manuscript "The opposing effects of hedonic and eudaimonic happiness on gene expression is correlated noise" #2016:03:9774:1:0:NEW). In addition to addressing the reviewer's concerns, I have followed the suggestion of the editor and modified the analyses and manuscript to make it more general. In particular, I have expanded the Monte Carlo simulation to investigate errors of all the inferential tests used, which both makes the results more general and addresses concerns of Reviewer 2. Additional modifications to the analyses include:

1. the modification of O'Brien's t-test to use the t-distribution instead of a null-distribution based on permutation of the data. The results are the same to the precision of a permutation test.
2. the modification of the permutation test to using Anderson's r_F^2 value as the test statistic. The results are the same to the precision of a permutation test.
3. I have added the GlobalAncova permutation F test as it is easily available in a R package, which readers will like
4. I have added a modification of the GlobalAncova permutation F-test to avoid violating the assumption of exchangeability.

The Monte-Carlo results for the F tests show excellent relative performance and are especially exciting for the field.

These changes have resulted in extensive modifications to the text. I did not revise with any kind of "track changes". I did create a diff.tex file (differences in to LaTeX files) I could not get this file to compile. I have uploaded a blank track changes file so that I could submit the revision and hope this is okay.

I am excited to submit these revisions and believe it ready for publication. My specific responses to the reviewers are attached below.

Sincerely,

Jeffrey A. Walker

Response to Reviewers

Response to comments from reviewer 1

Basic reporting

Very well written paper. Well structured. Although maybe a bit long: the arguments that the FRED2015 analysis is faulty would still be highly convincing with only half the evidence now presented by the authors.

I could not find the sample size of the FRED2013 paper. It would be helpful to have that mentioned.

Response: Thank you for the positive feedback. The sample size for FRED2013 was given in the first paragraph of METHODS so I assume this slipped past the reviewer. I have left this text unchanged.

The analyses and results are clear and convincing. Perhaps it could be mentioned that the ROAST analysis is small sample exact, whereas the permutation approach of Freedman and Lane is only asymptotically valid.

Response: I have added this suggestion to the methods section under the ROAST description

Response to comments from reviewer 2

Basic reporting

The author of the current article concedes that he lacks expertise in psychology, and a Google Scholar search suggests that he has not published in the area of genomics either. Nevertheless, he proceeds to reanalyze the data from Fredrickson et al. (2013) and one of the two studies presented in Fredrickson et al. (2015).

Response: the focus of my manuscript is the statistical analysis of data with multiple outcomes. While I am a biologist, my training is in multivariate statistics and my CV supports this.

The article reports the results of four non-parametric methods. A weakness is that the author proceeds without justifying why non-parametric approaches are superior to the parametric approaches used in Fredrickson et al. (2013, 2015).

Response: I agree that the multiple criticisms of the GLS method used by Fredrickson is not sufficient so I now (1) highlight the multiple concerns raised by the authors own analysis, 2) add citations to support my claim that of downward biased standard errors in the mixed model, and 3) extend the simulation of error types to all of the inferential methods.

The most significant issue involves PeerJ editorial criterion 2, as this paper does not 'describe original primary research.' All data analyzed here appear to have been published previously elsewhere, and the present manuscript reports only additional secondary analyses of those existing data. The present analyses appear to represent a step backward in terms of analytic rigor, so they would not seem to constitute any distinctive methodological contribution either.

Response: The Monte-Carlo simulations is the primary contribution and in the revision, this is greatly extended. I don't know what the reviewer means by "secondary analysis" other than an attempt to trivialize the results. The other two reviewers disagree that the present analysis represents a step backward in analytic rigor.

he never provides any clear reason to doubt the validity of the Fredrickson et al. (2015) mixed model results (and my read below is that they are more sound and appropriate than the new results reported here)

Response: In the paragraph starting at line 281, I give a list of five reasons (based on various model checks) to infer very unstable coefficients from the GLS model used by Fredrickson et al. 2015. To this I have added the increased (doubling) standard error of the coefficients estimated via the bootstrap, which is consistent with other simulations that I cite. In addition, I note the authors own contradictory results, most of which are buried in the supplement and not discussed. This includes 1) the conspicuous absence of the GLS model applied to the 2013 data and 2) the analysis of the 2015 data with an unstructured error matrix (FRED2015 Supplement Table 3) that contradicts the analysis using the compound symmetry matrix! Regardless, the reviewer makes no attempt to criticize any of these reasons for skepticism of the 2015 results, but only states "no clear reason" was given to doubt the prior results.

I have also added citation to Littell, R.C., Stroup, W.W., Milliken, G.A., Wolfinger, R.D., and Schabenberger, O. (2006). SAS for mixed models (SAS institute), that is useful in this context since FRED15 provided the SAS script in their supplement but they did not used the KR correction. Here is what Littel et al state:

"*Also, PROC MIXED computes so-called naive standard errors and test statistics: it uses estimated covariance parameters in formulas that assume these quantities are known. Kackar and Harville (1984) showed that using estimated covariance parameters in this way results in test statistics that are biased upward and standard errors that are biased downward, for all cases except independent errors models with balanced data. Kenward and Roger (1997) obtained a correction for standard errors and F-statistics and a generalized procedure to obtain degrees of freedom. The Kenward-Roger (KR) correction is applicable to most covariance structures available in PROC MIXED, including all of those used in repeated measures analysis. The KR correction was added as an option with the SAS 8.0 version of PROC MIXED and is strongly recommended whenever MIXED is used for repeated measures. Guerin and Stroup (2000) compared Type I error rates for default versus KR-adjusted test statistics. Their results supported Kenward and Roger?s early work: unless you use the adjustment, Type I error rates tend to be highly inflated, especially for more complex covariance structures.*"

The author concludes that the multiple non-significant findings from his analyses indicate the absence of a true association between well-being and the gene expression profile. However, the non-significant findings may be equally well explained by the substitution of less efficient non-parametric statistical analyses for the Fredrickson et al. (2015) mixed model analysis.

Response part 1: I thought nowhere did I imply that the association was zero but simply biologically "trivial" or statistically "undetectable" (the language that I try to use). I don't doubt there is an association. There HAS to be some association between expression level and happiness because both are necessarily related to real physiology. I do believe this association is too trivially small to be biologically meaningful and the coefficients are too poorly estimated to estimate their magnitude or even sign with any precision. I have

modified the manuscript throughout to clarify this.

Response part 2: The reviewer continues to be focused on efficiency at the expense of the two well know issues with linear mixed models: upward biased coefficients and downward biased standard errors. I now add a monte carlo simulation of type I error for all methods and not just the GLS method and show the inflated errors in GLS but not the non-parametric methods (and the parametric O'Briens).

This problem is compounded by potential biases arising from this author?s apparent specification of an exchangeable covariance structure for his analyses. The exchangeability assumption is contradicted by marked heterogeneity in both variance and covariance across the multiple genes analyzed (which both Fredrickson et al. and this author note, but only the former address by using more appropriate error structure specifications).

Response: It's not clear to me to what the reviewer is referring. The permutation procedure relies on exchangeability between subjects but not within, so the non-exchangeable error covariance structure is preserved in the Freedman and Lane permutation. By contrast, FRED2015 did not report results from a model allowing heterogenous (unstructured) covariances in the main text. They do in the supplement and the results contradict the main paper results and those of 2013. This is conveniently ignored by the reviewer and by Fredrickson et al. but I now raise this in my manuscript. I also now add a GEE test with the within-subject exchangeable assumption but both the literature and my monte carlo simulation of error rates show this is superior to the GLS

Given that this paper uses less powerful and less appropriate statistical models, is no surprise that it yields different and weaker results. The non-significant results observed here may reflect more about the data analyst?s choice of less efficient and accurate analytic methods rather than anything about the substantive data. Loss of power and precision would be observed in any analysis that substituted nonparametric techniques for parametric models when the latter are appropriate.

Response: Again, this criticism assumes that the GLS model is behaving well and all the model checking that I report shows that it is not.

This article also fails PeerJ editorial criterion 2 in providing insufficient information about the exact procedures performed. At a minimum, this article needs to include a transcript of all statistical code (e.g., in appendix or supporting information file) so readers can better understand exactly what has been done, how it differs from other analyses, and why. (github is not sufficient because those postings often disappear)

Response: The code was made available to the reviewers. Github is a very common respository for published code from, for example PLOS Computational Biology. However, I am happy to deposit the code at some site where I cannot delete the file.

Concerns about technical rigor/statistical soundness are also raised by multiple presentational errors or misrepresentations, detailed below. For example, the current author (1) characterizes the Fredrickson et al. (2015) analysis as a 'GLS' 'marginal model' when in fact Fredrickson et al. (2015) report using a standard conditional mixed linear model estimated by maximum likelihood;

Response: I refer to it as a GLS because that is exactly what the authors used if one looks at the supplement. The name for the analysis varies among sources but I think the best name for it is a "fixed-effects linear model with correlated error", which *is* a linear

mixed model but one in which the random effects are not explicitly modeled (the random effects are implicitly modeled in the specified error matrix). The reviewer uses the term "conditional mixed linear model" which is usually used in the context of linear mixed models in which the fixed effects are conditional on explicitly modeled random effects. Fredrickson et al did NOT analyze their data as a conditional linear mixed model as suggested by the reviewer. In the supplement to FRED15, the authors show the model, which is exactly my equation 4, which is a fixed effects model with correlated error. The authors also provide the SAS script in the supplement, which shows they used a repeated measures model using PROC MIXED. There is no random effect in this SAS model. According to Littell et al. ("SAS for mixed models. SAS Institute." Inc., Cary, NC (2006): 814.), SAS PROC MIXED uses GLS to estimate the fixed effects. And I was able to recover their results using the GLS function in R (as stated in the manuscript). By "marginal model" I mean FRED15 are only interested in the population averaged effect of hedonia and eudaimonia on expression levels and treat the estimated error as a nuissance parameter necessary to estimate the population-averaged (marginal) effects. It's not clear to me that the authors know what model they were using and I didn't want to imply this in my text but I try to make this clearer in my revision.

(2) misrepresents the number of data points analyzed as n = 122 or 198 rather than the 6,344 and 10,372 observations actually analyzed (which provide ample data for estimating the mixed effect models, contra the author?s claim);

Response: The reviewer knows that there are not "6344 and 10372" independent observations! My wording clearly states that these numbers refer to the number of subjects. I note here that the "effective" sample size is nowhere near 6344 and 10372 because of the correlated errors. In early versions of the manuscript, I cited work by Faes (Faes et al. "The effective sample size and an alternative small-sample degrees-of-freedom method." The American Statistician 63.4 (2009): 389-399.) but ultimately this computation isn't relevant other than for statistical learning

(3) asserts that marginal effects are 'typically' estimated by GEE, even though mixed model analyses are also widely used for this purpose, perfectly valid for it, and often more powerful (particularly in the present case involving mis-specified error structure for the GEE);

Response: Exhaustive searching of the relevant literature using Google Scholar failed to find more than a handful of analyses of multiple outcomes using a GLS model and no gene set associations using the GLS model, other than those by the authors. By contrast, Fredrickson et al. have completely ignored the multiple outcomes literature that developed within clinical medicine in the 90s and burst forward within gene set analysis in the 2000s.

(4) asserts that Fredrickson et al. 2013 and 2015 failed to test the 'delta' hypothesis of the difference (eudaimonic ? hedonic) when I was able to find those results easily in both cases;

Response: The reviewer is correct and I have modified my text accordingly.

(5) states that he?s reproduced the Fredrickson et al. analysis results when differences are apparent (i.e., in his Table 2);

Response: I did specify the model for the combined data slightly differently than Fredrickson 2015 (but not the individual datasets, which were replicated very precisely) because I left out the regressor "study". I have now added this and completely recover the

2015 results. This has trivial effects on my results. Also, I note that Fredrickson et al have corrected their 2013 analysis (PLOS 2016) so that their results match mine!

(6) asserts that the difference in his Table 2 results vs. Fredrickson et al. (2015) is due to variations in the posted data (I verified that it was not, and speculate that the difference instead arose from variations in model specification or estimation, such as use of GLS instead of maximum likelihood, compound symmetry rather than more appropriate error structures, etc.);

Response: This has nothing to do with maximum likelihood (which I used but didn't state in text) and compound symmetry (which I used BUT SO DID FREDRICKSON at least in their main text. Again, they report the results of the "more appropriate" unstructured error analysis in the supplement and these results contradict the main paper results an the 2013 results). See (5) above.

(7) concludes the difference between bootstrap SE?s and parametric model SE?s stems from models? 'sensitivity' to sampling variability (rather than to the nonparametric nature of the estimation algorithm and the well-known problem of bootstrap dilation (e.g., Bradley Efron (2010) Correlated z-Values and the Accuracy of Large-Scale Statistical Estimates, Journal of the American Statistical Association, 105:491);

Response: Efron's is an interesting results that I was unaware of but Efron's finding is for sparse, large p (number of responses), small N (number of subjects) datasets, which the reviewer fails to state. The reviewer also fails to cite the many papers, including the major text for SAS linear mixed models, showing the downward biased standard errors in the linear mixed model (which I now cite).

(8) in analyzing Type I error in the GLS method, produces Monte Carlo simulations that don?t correspond to the characteristics of the original analysis they are subsequently used to critique;

Response: It's unclear what is meant by "don't correspond to the characteristics of the original analysis". I used an X matrix with the same distribution (variances and correlations) as the data. A Y matrix with the same distribution as the data. The only difference was I simulated data in which there was no effect of any of the X on the Y. And I specified the model exactly as in the main paper of FRED2015 (GLS with heterogenous compound symmetry error variance).

(9) states that 'Random noise creates negatively correlated error' (which may be true for his Monte Carlo data, but is a nonsensical statement for empirical data analyzed by a different model; moreover all sources of parameter correlation seem to have been appropriately accounted for in the SE estimates for the mixed model).

Response: The empirical data analyzed by FRED15 was analyzed by GLS (multiple) regression so the phenomenon that I described (both verbally and mathematically) applies to FRED2015. This is a phenomenon due to a high correlation between two independent variables, not due to correlation among the outcome variables, as I clearly explained in the text but have added clarification in several points of the revised manuscript. Linear mixed models which use a GLS estimation (a generalization of OLS) does not account for it (which I show mathematically).

More misrepresentations could be cited. These inaccuracies left it difficult for me to trust any of the remaining assertions without verifying them myself. In the several cases where I did so, I noted a pattern of selective reporting and interpretation. For example, when I attempted

to reproduce the author?s preferred GEE analyses, I found residuals showing much greater deviations from normality and poorer fit than those of the Fredrickson et al. (2015) mixed effect models. In general, these analyses appear to move us farther away from analytic validity."

Response: I don't have a dog in this fight. There is no selective reporting (the irony here is thick). More importantly, I did not report GEE results (but I do in the revision). I did the exact same analysis as FRED2015. In response to reviewer 3, I now included the residuals vs. fitted values plot for the GLS analysis and add this plot for a GEE analysis and show the GEE is well behaved.

Validity of the findings Again, from my statistical colleague:

"The contribution fails PeerJ editorial criterion 3 by asserting substantive conclusions well beyond what could possibly be supported by the analytic results. Perhaps the most significant interpretive error is the author?s overall conclusion that the pattern of gene expression is 'simply correlated noise.' This conclusion commits the elementary statistical error of accepting the null hypothesis (see Goodman S. A dirty dozen: twelve p-value misconceptions. Semin Hematol. 2008 Jul;45(3):135-40. PubMed PMID: 18582619). If the author didn?t generate the data (e.g., by simulation) then it is impossible to say how the associations actually arose. At minimum the conclusions need to comply with the recent ASA consensus statement on appropriate interpretation of p-values (http://amstat.tandfonline.com/doi/pdf/10.1080/00031305.2016.1154108).

Response: I am well aware of the p-value misconception. This is why I used statements like this in the results: "the procedural bootstrap, the permutation t-test, the permutation O?Brien?s t-test, and the rotation z-test are very consistent and all fail to reject the null".

Response 2: Nowhere do I claim there is "no effect" of hedonia or eudaimonia on expression (the misconception). What I do claim is that the apparent pattern of opposite effects of hedonia and eudaimonia is a function of the correlated error. I claim this because I show mathematically and via simulation that the error is correlated and if the signal is small enough then the coefficients will tend to have opposite signs. This result is not based on a p-value. There may also (in addition) be a opposing effect of hedonia and eudaimonia on expression levels but the signal is not big enough to detect above and beyond the correlated noise. I have clarified this in the discussion and rephrased my conclusion as "The apparent replication of coefficients of opposite sign for $Hedonia$ and $Eudaimonia$ is consistent with trivially small effects (effectively equal to zero) in combination with the high empirical correlation between hedonic and eudaimonic scores (0.80 in FRED13 and 0.74 in FRED15). " and "The most parsimonious explanation of the apparent replication of opposing effects of hedonic and eudaimonic scores on CTRA gene expression is correlated noise arising from the geometry of multiple regression."

The most the author could validly claim is that the magnitude of associations estimated in his specific nonparametric analysis of the Fredrickson et al. data are not inconsistent with correlated noise as specified in that model.

Response: see my reply above

To meet the criterion of reporting all relevant results, the Discussion needs to note that the data are clearly inconsistent with correlated noise under the mixed model specification applied by Fredrickson et al. (2015). An accurate interpretation of the present results would also note that the non-significant results observed here may result from the comparative inefficiency

of nonparametric techniques used here, rather than from the absence of a true relationship in the population data. The ultimate question is whether the GEE or mixed model should be preferred on epistemological or goodness-of-fit grounds. Given the inconsistency of current nonparametric results with those of Fredrickson et al. (2013, 2015), it is also inaccurate to claim that results are 'unambiguous.' The primary conclusion is inaccurate on multiple grounds and greatly over-reaches what could ever be known from any secondary analysis of empirical data."

Response. I have addressed these comments above. In addition to the responses above, I note that the statistical colleague 1) selectively mines the statistical literature in favor of linear mixed models for inference on fixed effects, including failing to note any of the well known literature showing the inflated coefficients and depressed standard errors, and 2) does not address *any* of the tests I used to estimate p-values for eudaimonic or hedonic effects other than the very vague "non-parametric tests are less efficient than parametric tests", which may or may not be true with resampling/permutation tests as opposed to rank based non-parametric tests.

Response to comments from reviewer 3

Reviewer 3 (Anonymous)

Basic reporting

The article is clear and understandable. Experimental design This is an interesting study that uses a variety of techniques to investigate the relationships between CTRA gene expression and measures of happiness. Overall, the statistical work seems appropriate to me, and extends and applies mostly well-known statistical techniques ? namely a number of permutation and bootstrap approaches. The first bootstraps data rows, the second is a residual permutation procedure to generate a null (but retain the ability to use covariates), the third is similar with a different test statistic, and the fourth involves a rotation test, using a procedure that I am less familiar with. The fifth and sixth are perhaps the most critical ? they use the same procedure as Frederickson et al. (2015) but with bootstrapped, or permuted, data, and so test their procedure fairly directly. All of these approaches share a similar driving motivation, which is to test the robustness of the procedures used in two previous published papers, one of which attracted some previous critical responses, and the resulting conclusion. Alongside these analyses of the real data, there is also a more limited analysis of simulated gene-expression like data, to illustrate the fact that the GLS procedure can inflate type I error, when untrue modelling assumptions are made (particularly regarding the covariance matrix S_Y). Validity of the findings Overall I have no serious arguments with the methods used or the conclusions reached.

The bootstrap and permutation approaches both imply that GLS estimates of errors are substantively underestimated in these data, and correcting for this completely alters the conclusions in terms of result significance. Because individuals are bootstrapped wholesale, the (unknown) covariance structure of the data is conserved. The principle aim here is to estimate uncertainties of parameter estimates, using an approach robust to likely causes of model mis-specification.

Response: Thanks for the positive feedback!

Comments for the Author

I do have some specific points:

1) The null of no effect in either happiness measure, $\beta_{hedonia} = \beta_{Eudaimonia} = 0$ is a natural

one for a non-specialist to consider as a start point ? can the author explain why this is not of interest, to aid a general reader - or simply include this test and comment briefly? I believe the previous studies also investigated this question, via an F-test. This is natural as the authors say their aim is to study ?what is the evidence for effects of hedonic and eudaimonic happiness scores on CTRA gene expression?. Then separately setting $\beta_{hedonia} = 0$ and $\beta_{Eudaimonia} = 0$ (or indeed the same) become natural next questions, to evaluate whether one can separate the two coefficients.

Response - One issue that I worry about in my own manuscript and this comment is the extreme focus on "p-values" at the expense of the estimates and their errors, especially in a study with zero theory predicting causal pathways between the regressors of interest and the outcomes. Consequently, I'm just not interested in the general linear hypothesis $\beta_{hedonia} = \beta_{Eudaimonia} = 0$, especially given the extremely high correlation between Hedonic and Eudaimonic scores. As I now state in the paper, I have no doubt that there is some correlation between these variables - there has to be because they have to causally related by the underlying physiology. But the correlation is probably 1) trivially small and 2) extremely conditional on other variables. I add this view to my revision. Additionally, the revised manuscript is far more focused on the test performance and not the detailed results

2) In the abstract: ?is simply correlated noise?. It is not possible to be statistically certain of this, given only a failure to reject the null hypothesis. It would be better to write ?is consistent with correlated noise? or similar.

Response: I agree and now change the language to: "The apparently replicated pattern of gene expression is most parsimoniously explained as 'correlated noise' due to the geometry of multiple regression"

3) ?This small sample per parameter ratio is likely to result in overfit models, which, in turn, will result in unstable and inflated coefficients 78 (Harrell, 2015).? This seems a somewhat unjustified statement to me, because in fact there are 122/198 observations per gene. For each gene, only a small number of parameters must be estimated. The deeper problem raised here is the potentially highly complex correlation structure in the underlying data. This is a serious potential confounder, and would (should) result in a very large number of parameters to fit relative to the data size, resulting in the previous studies using what are likely over-simplified, rather than overly complex, models.

Response - I have removed this section and concentrate on the biased standard errors although there clearly is a problem with the coefficient estimation itself.

4) Following on from (3), I note that in the recent Fredrickson et al. study (2015), their Table S3 tests the same hypotheses under a general covariance structure, with a mixed effect linear model. This does not seem to be discussed in the Walker manuscript, but should be critically discussed ? because the manuscript currently implies that the compound symmetry covariance structure was the only one used. In the Fredrickson study their results for several measures change substantively when using the more general model, although they suggest convergence issues might be at fault.

Response - yes I was more focussed on not "beating up" the authors of the original paper but now include this in my revision as it is too critical to leave out.

5) ?Finally, and most importantly, the GLS model gains power by assuming that the

estimated regression coefficient is a common effect for all genes?. The current wording reads as if this point potentially invalidates the analysis of Frederickson et al. ? because the common effect assumption is likely not to be true. However, what is important for the argument of the present study is whether this invalidates the test (rather than altering its power ? indeed increasing the power is a favourable outcome if true). Because the null hypothesis is that all effect sizes are (equal and) zero, an alternative of equal non-zero effect sizes results in a valid test in principle - so I think this criticism should be withdrawn. Indeed, if effect sizes are heterogeneous this would reduce the power, but not increase the type I error rate, of the GLS test. Again, the point here is really that the previous authors used an overly simplified model, therefore not robust to departures from modelling assumptions.

Response - Agreed and I have removed this.

6) In the recent Fredrickson et al. study, I believe they do actually test for a difference in coefficient of Hedonia or Eudaimonia (at least in a sense). E.g. from their paper: ?In direct comparison of standardized association coefficients, the magnitude of CTRA association with eudaimonic scores significantly exceeded that for hedonic scores (t(104) = -2.59, p =. 0109).? This seems to be a t-test for a difference in coefficients, but is described here as a test of p-values ? language which I think is misleading and should be explained more clearly in terms of its difference with the procedure here - i.e. the coefficients are standardised before comparison in the Fredrickson et al. case, but not in the Walker case, as I read it.

Response - I have modified my text to acknowledge this.

7) On p3 ?(ŕ?hedonia and ŕ Bhedonia) is a typo (repeated).

Response - fixed

8) It would be helpful to include evidence of model mis-specification that underlies the identified problems with the GLS approach here - for example the behaviour of the observed residuals (discussed briefly) could be shown, or tested relative to their expectations.

Response - Agreed so I have added the figure that shows the diagnostic residual vs. fitted values for the GLS and also added this for the GEE estimates.

Response to comments from the Editor