For the Editor:

A brief note, to expand on some of my comments to the authors:

The paper purports to provide a “detailed study of potential instrumental systematic contamination” to the B-mode signal detected by the Bicep2 collaboration. While it contains a great deal of valuable information regarding the data and the instrument, it does not include a description of the instrument and analysis that is sufficiently comprehensive to fully assess the role non-astrophysical contributions to the B-mode power reported in the Results Paper.

The paper is primarily a (valuable) study of certain aspects of the signal estimation and the treatment of beam effects. As such, the scope is more narrow than the typical ’systematics paper’ in this field, for example those of QUIET and WMAP

http://iopscience.iop.org/0004-637X/768/1/9

http://adsabs.harvard.edu/abs/2011ApJS..192...14J

http://adsabs.harvard.edu/abs/2007ApJS..170..263J

The paper makes reference to a corresponding ‘Beams Paper’. Given the overwhelming focus of this paper on beam and pointing effects, one wonders why this paper is distinct.

Besides the comments I include below, there are three topics in particular, related to the role of systematics in the Results Paper that are not adequately addressed in the current text.

1. The role and significance of the Jackknife failures
2. A detailed description of the noise model used in both the determination of the auto-spectra and the null tests, in particular the role of correlations in the effective noise
3. Resolution of the apparent systematic discrepancy between the B2 auto-spectra and the B2/Keck cross-spectra in Figure 9 of the Results Paper

The discussion present in the current draft is worthy of publication as a companion to the Results Paper, but the above items must be addressed before publication as a comprehensive discussion of systematics in the Bicep2 experiment.

Paper Review:

The Bicep2 experiment has produced high signal-to-noise measurements of polarized emission in a small (<1% of the full sky) patch of the sky at high Galactic latitude. This emission includes a dominant contribution from the even parity (E-mode) polarization expected from the CMB in the context of the standard LCDM cosmology, as well as a significant contribution from weak lensing of the CMB polarization and from thermal emission from Galactic dust. The latter contribute to both the even- and odd-parities of polarization (E- and B-mode).

This manuscript provides a thoughtful and well written description of how the analysis has handled certain beam effects, pointing, and a particular subset of systematic effects in the Bicep2 data analysis. The interpretation of the Results Paper is difficult without a thorough treatment of these topics, making the publication of a companion paper including this supplementary information a priority.

The reviewer acknowledges the enormous effort that is required to provide a comprehensive study of instrumental systematics - in many ways more daunting than the cosmological analysis. The authors should be commended for their efforts in this regard, the further comments not withstanding.

There are a number of substantial issues that require further attention. These are briefly summarized before running through section by section.

I) General comments regarding substantial issues

a) Noise estimation:

The Results Paper relies on the accurate calculation of auto-power spectra, which requires extremely accurate determinations of the statistical properties of the noise to correct for the additive bias. The difficulty in determining the noise properties at the required level of accuracy has driven the field to shift instead to methods using cross spectra of data subsets that have uncorrelated noise contributions.

There is insufficient detail regarding the determination and qualification of the noise model to assess the importance of residual additive biases in the power spectra. The suite of jackknife tests provides some measure of this, modulo the potential importance of correlated noise, but again the description of the noise model is insufficient to judge. A very detailed description of the noise model applied both to the full data set and the jackknives, including a discussion of the potential role of noise correlations, is necessary, and missing from the current manuscript. Alternatively, a selection of cross-spectral results showing consistency with the auto-spectra would represent a robust check of fidelity of the noise model, modulo certain mechanisms of correlated noise.

The Results Paper is not limited by statistical noise, so there is no reason not to trade a bit of that statistical power for robustness against noise correlations and mis-estimation of the noise.

b) Jackknife tests:

The discussion of the jackknife tests needs to include a measure of the statistical weight, so that there reader can assess the degree to which each limits the significance of systematic contamination in the CMB spectra.

The discussion at the end of Section 7 dismisses the low PTEs as being not unlikely, with the exception of the several EE failures which the authors suggest result from pair-to-pair CMB gain variations. Table 1 of the Results Paper shows that the low PTEs observed both in the \chi^2 and \chi tests are bunched, making them rather unlikely for example in the Tile EE, Alt-deck EE, Mx row EE/BB/EB, Tile and deck EE/EB and Focal Plane inner/outer EE/BB/EB. More discussion of these is warranted, and again the amplitude of the failures should be discussed in the context of the statistical noise and the signal level in the summed spectra.

B2/Keck obtains the relative calibrations of detectors using signal from periodic sky dips. While this allows for the most efficient removal of common mode atmospheric pickup, the difference in the spectral energy distribution of the atmosphere (which is also time variable) and the CMB ensures that the differential gain to the CMB is mismatched at some level (and perhaps not stable over relevant timescales). This ensures that some level of CMB temperature signal will contribute to the pair differenced data products. This leakage transforms in polarization in the same way as statistical noise, and therefore represents an irreducible level of correlated noise even if there are no other mechanisms that produce it.

Figures 3 and 4 suggest that this level is significant, and therefore further discussion is warranted.

Figure 9 of the Results Paper suggests statistically significant systematic differences between the B2 and Keck data. This needs to be addressed somewhere in the paper.

c) Polarized side lobe pickup:

In all of the discussion surrounding the beams, there is no discussion on the role of polarized side lobe response. Perhaps this has been covered sufficiently in other publications; if so these should be referenced. At these low levels there are extremely stringent requirements on the polarized pickup from the Galaxy and from the ground, although the latter should be somewhat mitigated by the template removal (at fixed boresight angle).

II) Comments, Section by section:

1. Introduction:

First paragraph, “with the addition of very few additional cosmological parameters”. What additional parameters? Planck has shown that none but the usual six, determined from the temperature power spectrum, are required to reproduce the observed even-parity polarization modes. The converse is true as well, and of course the temperature and polarization anisotropies contain complementary information regarding those parameters, essentially because of the expanded set of modes that are observable.

Third paragraph. This paper presents extensive studies of specific classes of possible systematics, and these are found to contribute to the b-mode signal at a negligible level. However, the classes investigated are not exhaustive in the paper as it stands.

2.1 Instrument design

Fourth paragraph, last sentence. It is stated that the contribution of spurious signals from temperature drift are constrained, but there is no quantitative description. The discussion should be made quantitative or be removed.

2.2 Observational strategy

First paragraph, "… during which time the sky rotated …". I understand what is meant, but in this context "sky rotation" is a term typically applied to the variation in position angle, which would have very different implications for the data. It would be better to use some other phrasing here.

2.3 Summary

Time domain filtering. How is atmospheric noise accounted for in the noise model? The polynomial filtering is applied to the pair differenced data and "such a filter also removes any large angular scale contamination that might not average down." If the atmosphere is unpolarized, what is the origin of the observed residual? A failure of the relative gain procedure? Non-stationarity in the gain on the relevant timescales? If these, what are the implications for the subsequent analysis?

3. Beam systematics

The classification of centroid offsets as a systematic is confusing - it not so much a systematic as it is an inadequacy of the pointing model used in the signal estimation pipeline. Why wasn't the pointing model updated with the knowledge of the beam centroids?

In all of the subsequent discussion, and in that of the Results Paper, it appears that this differential pointing is by far the dominant effect. There is a great deal of discussion regarding the higher order beam effects, but these are shown to be negligible. While this a valuable result, it would seem like the discussion is disproportionate to its importance in the final analysis.

Table 1: The "Monopole" refers more often to differences in the dc level between two datasets, rather than a differential forward gain, as is meant here. Gain would be a better term. Although I suspect it is irrelevant, it would be good to quantify the degree to which true monopole levels affect the analysis (degenerate here with coherent

rotation).

5. “External” beam measurements

Paragraph 3. In general, the flat mirror will have an effect on the A/B differences, both in the beam and the polarimetric calibration. In order to understand the correspondence between these data and the beams on the sky, a more detailed discussion of the optics is required, including, for example, physical optics modeling of the near field flat.

7. Jackknife tests

General comments:
The null spectra are not universally sensitive to noise correlations (or other effects that are equivalent) that may potentially bias the summed data. This needs to be addressed here, or in a section dedicated to the noise modeling.

The pair-to-air variation in gain are invoked to explain the low PTE’s in EE. The BB data are not particularly low s/n either. These are claimed to explain the low PTEs, but there is no quantitative discussion of how the values estimated in Figure 7 propagate to the jackknife residuals.

9.3 Gain variation

What of temporal gain stability? There is no quantitative discussion of the amplitude of the gain variations measured by the sky dips, the expected amplitude of variations between dips, the need for gain model, the impact of the temporal variation in the relationship between the gain measured on the atmosphere and that of the CMB, etc. The section is important, but as written is remarkably terse.

10. Conclusions

The ‘unprecedented systematics control’ is indeed impressive, but not without precedent. If the issues described above are addressed, the demonstrated systematic control will be comparable to that of QUIET and others, which should be referenced as well.