

Quantifying the global parameter tensions between ACT, SPT, and *Planck*

Will Handley^{1,2,3,*} and Pablo Lemos^{4,†}

¹*Astrophysics Group, Cavendish Laboratory, J.J.Thomson Avenue,
Cambridge CB3 0HE, United Kingdom*

²*Kavli Institute for Cosmology, Madingley Road, Cambridge CB3 0HA, United Kingdom*

³*Gonville & Caius College, Trinity Street, Cambridge CB2 1TA, United Kingdom*[‡]

⁴*Department of Physics and Astronomy, University College London,
Gower Street, London WC1E 6BT, United Kingdom*

 (Received 23 July 2020; revised 22 December 2020; accepted 3 March 2021; published 26 March 2021)

The overall cosmological parameter tension between the Atacama Cosmology Telescope 2020 fourth data release (ACT) and *Planck* 2018 data within the concordance cosmological model is quantified using the Suspiciousness statistic to be 2.6σ . Between ACT and the South Pole Telescope (SPT) we find a tension of 2.4σ , and 2.8σ between ACT and *Planck* + SPT combined. While it is unclear whether the tension is caused by statistical fluctuations, systematic effects or new physics, caution should be exercised in combining these cosmic microwave background datasets in the context of the Λ CDM standard model of the universe.

DOI: [10.1103/PhysRevD.103.063529](https://doi.org/10.1103/PhysRevD.103.063529)

I. INTRODUCTION

As cosmological datasets increase in quantity and quality, so does our capacity to use them to pin down the properties of our universe [1]. The error bars on the measurements of cosmological parameters have narrowed over recent years and discrepancies between datasets (or “tensions”) have begun to emerge. While this is most stark when examining differing observations of the Hubble parameter between early and late time cosmological probes [2–4], other more minor tensions arguably exist in clustering parameters between weak lensing and the cosmic microwave background (CMB) [5,6] and in cosmic curvature between the CMB and CMB lensing/baryon acoustic oscillations [7–9].

When a substantial tension occurs, it may indicate either a systematic error in how either or both of the datasets have been gathered and analysed, or more excitingly may hint at evidence for new physics if extensions or modifications to our concordance model can bring the inferred parameters back into alignment.

In the case of the “Hubble tension” where a single obvious cosmological parameter such as the present day expansion rate H_0 is discrepant by $\sim 5\sigma$, there is little doubt that something is fundamentally wrong. The other tensions are more subtle, in that they are only visible in complicated combinations of the parameters. As shown by Fig. 1, in modern cosmology, error bars on the parameters of our

universe are represented by high-dimensional Bayesian probability distributions. Visualising a “distance” between these degrees of belief is challenging, and in recent years a good deal of theory has been developed for defining a variety of metrics of discrepancy [10,11].

The latest Atacama Cosmology Telescope (ACT) data release 4 [13–15] represents the most recently acquired CMB data, with two other measurements of the CMB power spectrum across a wide range of multipoles being provided by the *Planck* satellite [16,17], and the South Pole Telescope (SPT) [18]. By eye it is clear that in the ACT data some parameters such as the spectral tilt of the primordial power spectrum n_s are mildly discrepant, but it is always possible in a high dimensional parameter space that such discrepancies occur by chance and are unremarkable.

In this paper we discuss how this tension is quantified rigorously using the global Suspiciousness statistic [[19], henceforth H19], and find that ACT is in mild-to-moderate tension with *Planck* and SPT, at a similar or greater level to that found in weak lensing data. We place ACT’s own global tension analysis in the context of the tensions literature, and extend it by considering SPT data and further emphasize the perils of focusing too closely on lower-dimensional views onto the cosmological constraints.

II. METHODOLOGY

Quantifying tension between high dimensional posterior distributions is a non-trivial problem, even under the approximation of a Gaussian distribution. This has led to a large number of papers describing methods to quantify tension in high dimensional problems [for reviews, see

*wh260@mrao.cam.ac.uk

†pablo.lemos.18@ucl.ac.uk

‡<https://www.kicc.cam.ac.uk/directory/wh260>

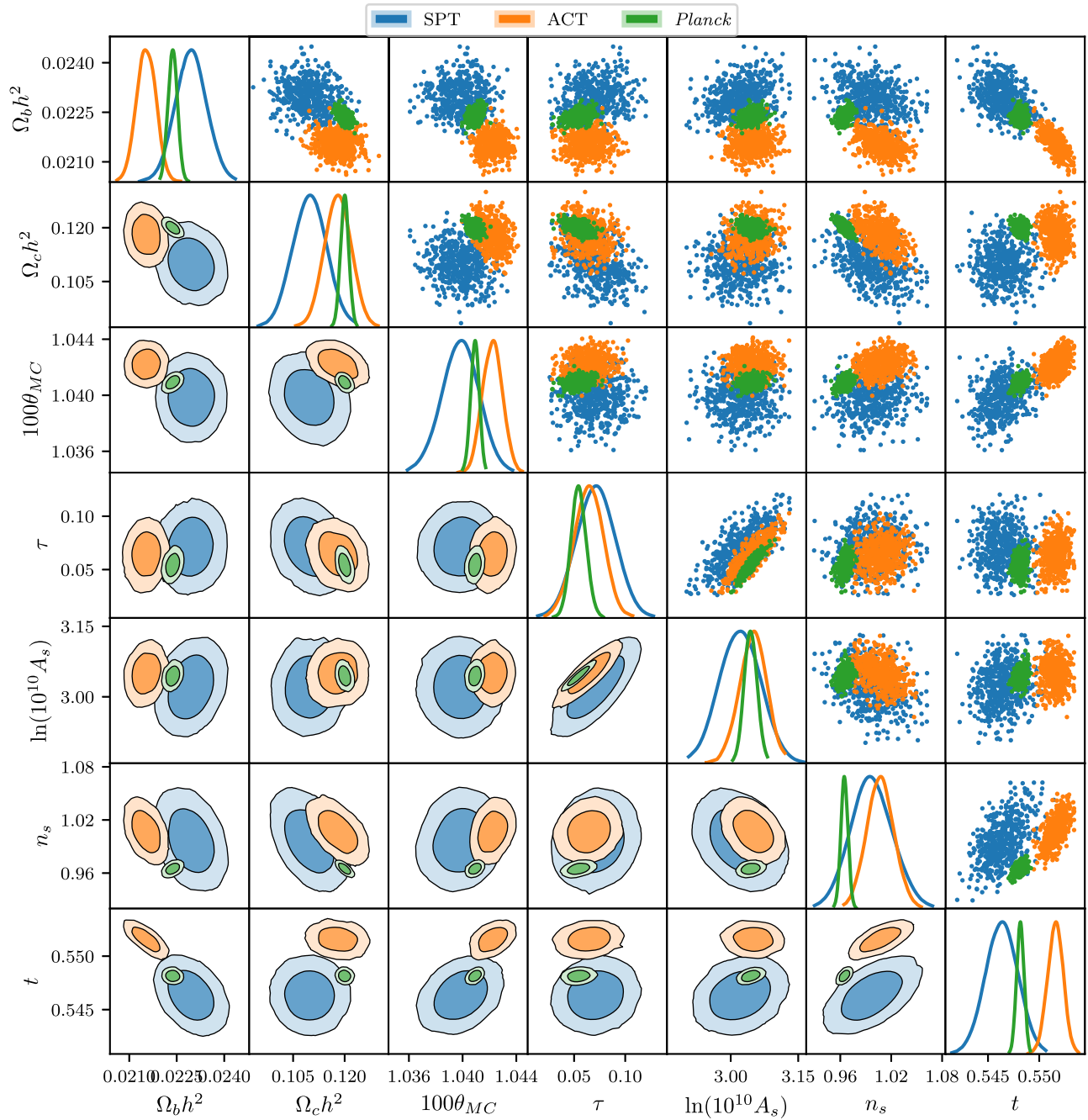


FIG. 1. Measurements of the six parameters of the concordance Λ CDM model using data from the South Pole Telescope (SPT, blue), the Atacama Cosmology Telescope (ACT, orange) and the *Planck* satellite (green). Plots along the diagonal show one-dimensional marginalized probability distributions normalized to equal height, below the diagonal show isoprobability contours containing 68% and 95% of the 2d marginal probability mass, and above the diagonal show samples drawn from the full probability distribution. Of the six cosmological parameters, visually ACT stands out in tension from the other two most clearly in the $n_s - \Omega_b h^2$ plane. We can artificially emphasize this further by computing and plotting the linear combination coordinate of maximum tension between ACT and *Planck* $t = -\Omega_b h^2 + 0.022\Omega_c h^2 + 34\theta_{MC} - 0.092\tau + 0.05 \ln(10^{10} A_s) + 0.067n_s$, which by construction will have a tension of $\chi = 4.15\sigma$. Marginalized plots can therefore overemphasize tension by ignoring the other active coordinates, but the headline statistics in Table I are derived from considering the entire distribution as a whole. Plot produced under ANESTHETIC [12]

[10,11]]. Working in a Bayesian framework, as most cosmological analyses do, arguably the most natural way to quantify tension is using the Bayes Ratio [20], defined as

the ratio of the probability that the two datasets are described by a single set of parameters, to the probability that they are described by separate sets of parameters

$$R = \frac{P(A, B)}{P(A)P(B)} = \frac{\mathcal{Z}_{AB}}{\mathcal{Z}_A \mathcal{Z}_B}, \quad (1)$$

where P represent a probability, we have omitted the dependence of both probabilities on an underlying model, such as Λ CDM, and \mathcal{Z} is the Bayesian Evidence. Furthermore, we have assumed that both data sets are independent, an assumption that we further comment on later. High values of R correspond to concordance, and low values are indicative of discordance, with R often interpreted on a Jeffreys' scale [21,22]. The main issue of this tension metric, in particular for the analysis of cosmological data sets, is that it is easily proven that R is proportional to the prior volume of shared parameters. Therefore, R cannot be used for analyses that use deliberately flat and wide uninformative priors, such as the analyses of *Planck*, Dark Energy Survey [DES, [22]], Kilo Degree Survey [KiDS, [6]], ACT, SPT, etc. without the arbitrary width of this prior affecting tension assessment. A more detailed interpretation of this discussion can be found in H19.

Motivated by this, H19 defined a new statistic, the *Suspiciousness* which keeps all the desired properties of Eq. (1), but corrects for this undesired dependence on the prior volume. To do so, we divide the Bayes ratio in two components: Information and Suspiciousness. The Information is defined as:

$$\log I = \mathcal{D}_A + \mathcal{D}_B - \mathcal{D}_{AB} \quad (2)$$

where \mathcal{D} is the Kullback-Leibler divergence [23]. The Information contains the dependence on the prior volume, therefore by removing it, we obtain a statistic that does not depend on it, but is composed of well-defined Bayesian and information theoretic quantities and is therefore covariantly insensitive to reparametrization of the space. Therefore, we define the Suspiciousness as:

$$S = \frac{R}{I}. \quad (3)$$

In the language of priors, the Suspiciousness may be interpreted as the most cautious Bayes ratio R corresponding to the narrowest possible priors that do not significantly alter the shape of the posteriors [24].

A significant innovation to the field which we highlight here, first noted in the Appendix F.3 of [25] and explored in detail in [26] is that since $\log \mathcal{Z} = \langle \log \mathcal{L} \rangle_{\mathcal{P}} - \mathcal{D}$, the Suspiciousness can be computed from MCMC chains via

$$\log S = \langle \log \mathcal{L}_{AB} \rangle_{\mathcal{P}_{AB}} - \langle \log \mathcal{L}_A \rangle_{\mathcal{P}_A} - \langle \log \mathcal{L}_B \rangle_{\mathcal{P}_B}. \quad (4)$$

This observation means that so long as one has posterior samples for each of the datasets run separately and in combination, one may compute the Suspiciousness without explicitly computing the Bayesian evidence. However, it

should be noted that in non-CMB applications only a portion of the parameters are constrained, resulting in hypersurface-like posteriors which are extremely challenging for traditional posterior samplers, but present little challenge for nested samplers.

If the posteriors are such that we may approximate them in the cosmological parameters by a Gaussian (an approximation which is reasonably justified as shown by Fig. 1), as derived in H19, if the d -dimensional posterior distributions are Gaussian in the parameters with means and covariance μ and Σ , then the Suspiciousness is

$$\log S = \frac{d}{2} - \frac{\chi^2}{2}, \quad (5)$$

$$\chi^2 = (\mu_A - \mu_B)(\Sigma_A + \Sigma_B)^{-1}(\mu_A - \mu_B). \quad (6)$$

This may be turned into a tension probability via the survival function of the chi-squared distribution

$$p = \int_{\chi^2}^{\infty} \frac{x^{d/2-1} e^{-x/2}}{2^{d/2} \Gamma(d/2)} dx, \quad (7)$$

and calibrated using a σ -tension by analogy with the Gaussian case using the inverse of the complementary error function:

$$\sigma(p) = \sqrt{2} \text{erfc}^{-1}(1 - p). \quad (8)$$

Note that, while several methods to quantify tension have been proposed in recent years, they are often built to recover Eqs. (7) and (8) in the case of Gaussian posterior distributions. Therefore, if this work were performed using tension metrics such as Monte-Carlo Parameter Shifts [27], Parameter Shifts in Update Form [28], or EigenTension [29], we would expect to obtain very similar, if not the same results, under the Gaussian approximation used in this work. This is also equivalent to the multivariate measure of tension used in the ACT paper [13].

It should be noted that alternative measures of tension have also been defined and explored that are specialized for the case when two datasets are correlated [27,30]. In particular, [24] extended the formalism described in this section to the case of correlated data sets. Applying this to the case of CMB datasets such as *Planck*, ACT, SPT and WMAP (which are correlated by virtue of their measuring the same sky) will form the subject of a future paper.

III. DATA

In this work we analyze the three latest CMB data sets, *Planck*, SPT and ACT. As with all cosmological analyses, when considering combining or comparing them at the likelihood level we implicitly assume that the datasets are independent, even though this may not strictly true. Examining the effect of relaxing this assumption will form

the subject of future work. It should also be noted that the prior treatment for τ is different across the three collaborations, one of the aims of future work will be to treat this in a consistent manner for all three cases. The ACT analysis uses a CMB-derived prior for τ , so there is correlation between posteriors for the τ parameter. A more complete analysis could adjust the tension in either direction, since correlations in τ act to reduce the dimensionality to less than $d = 6$, increasing the tension [19,30,31], but since τ is a degeneracy-breaking parameter it can have dramatic effects in moving the relative locations of posteriors, increasing or reducing tension.

A. Planck

The *Planck* mission [32] was a space observatory that measured the CMB for four years between 2009 and 2013. *Planck* observed the sky in nine frequencies, between 30 and 857 GHz, with the goal of detecting both temperature and polarization anisotropies, and accurately removing foreground effects. *Planck* measured the power spectrum of temperature anisotropies in multipoles $\ell \in (2, 2508)$, and for E-mode polarization in multipoles $\ell \in (2, 1996)$, providing the most powerful constraints in the parameters of the Λ CDM cosmological model to date.

Beyond the already mentioned tensions in H_0 cosmic curvature, and with weak lensing; the most puzzling aspect of the *Planck* analysis is arguably the A_L parameter.¹ A_L was introduced for internal consistency checks [33], and can smooth the peak of the *Planck* power spectrum. *Planck* [32] reports a value $A_L = 1.180 \pm 0.065$ for the combination of temperature and polarization, meaning that the data seems to prefer more smoothing of the peaks than the best fit Λ CDM cosmology provides. While it has been discussed that this could be caused by a statistical fluctuations, especially since the significance is lower for different versions of the likelihood [34], it has also been hypothesized that it could be a hint of new physics [8], although no theoretical model that produces this effect exists in the literature. It is important to point out that, while this effect is similar to that of CMB lensing, *Planck* lensing measurements [35] are compatible with $A_L = 1$.

Throughout this paper we use the *Planck* legacy archive chains derived using the baseline TTTEEE + low ℓ + lowE + lensing, and have confirmed that our conclusions are insensitive to excluding the lensing portion of the likelihood.

B. South Pole Telescope

We make use of the South Pole Telescope measurements of temperature and polarization from the 500 square degree analysis of their SPTpol instrument [18]. This analysis used

data at 150 GHz to produce power spectra for the E-mode polarization (EE) and the temperature-E-mode cross-spectrum (TE). It should be noted that the TE and EE have been identified as being in disagreement, so strictly this internal combination should also be viewed with suspicion. The main advantage of SPTpol with respect to *Planck* is its higher resolution, which allows it to measure much smaller scales, covering a multipole range $\ell \in (50, 8000)$. However, because of its smaller sky coverage, SPTpol cannot obtain information on large scales, and as a consequence produces parameter constraints that are weaker than those from *Planck*. SPT [18] reports constraints that differ from *Planck*'s, in particular when only SPTpol's high multipoles are used, but the significance of this reported discrepancy is not quantified. *Planck* [32] used a parameter difference statistic, and found no evidence for statistical inconsistencies between the two analyses. Curiously, performing an A_L analysis on SPTpol yields a value lower than one, $A_L = 0.81 \pm 0.14$

C. Atacama Cosmology Telescope

Finally, we use the Atacama Cosmology Telescope (ACT) posterior samples² from Data Release 4 (DR4), which used 6000 square degrees at 98 and 150 GHz to produce power spectrum for temperature and polarization extending to $\ell = 4000$. Their results by eye appear to be in tension with *Planck*, and ACT [13] report a global tension with *Planck* consistent with that recovered in this paper.

IV. RESULTS

Our results are summarized in Table I and Fig. 1. When the Suspiciousness tension quantification techniques are applied to the ACT data products in comparison with the *Planck* baseline, we find a tension probability of $p = 0.86\%$, with a corresponding Gaussian-calibrated tension of 2.63σ . This level of discrepancy is generally termed mild-to-moderate, and is comparable with some of the larger tensions found between weak lensing and CMB data [H19, [6]].

The degree of discrepancy between *Planck* and ACT is consistent with the level of tension reported by ACT [13]. It is important to note that a global tension quantification such as Suspiciousness does not depend on any specific direction choice in parameter space, nor on the choice of parameters. It also naturally takes into account the effect that in having $d = 6$ parameters, it is not improbable that some would be in strong marginal tension by chance. One can make this point explicit by computing an artificial parameter t defined as the linear combination of the other parameters θ which maximizes tension. In the Gaussian case this ‘‘maximum tension parameter’’ may be computed as

$$t \propto (\mu_A - \mu_B)^T (\Sigma_A + \Sigma_B)^{-1} \theta, \quad (9)$$

¹Often known as ‘‘lensing parameter’’ or A_{lens} , but we will refrain but these names as we believe they can be misleading.

²phy-act1.princeton.edu/public/zatkins/ACTPol_lcdm_1.txt

TABLE I. Global tensions between CMB datasets. For each pairing of datasets we report the χ^2 value calculated using Eq. (6), the corresponding tension probability p from Eq. (7) that such datasets would be this discordant by (Bayesian) chance, a conversion into a Gaussian-equivalent tension using Eq. (8) and finally the Suspiciousness from Eq. (5). Addition signs in the left column indicate combining the datasets at the likelihood level, and combinations below the line should be viewed with suspicion on account of their discordance reported above the line.

Dataset combination	χ^2	p	Tension	$\log S$
ACT vs <i>Planck</i>	17.2	0.86%	2.63σ	-5.60
ACT vs SPT	15.4	1.77%	2.37σ	-4.68
<i>Planck</i> vs SPT	9.1	16.82%	1.38σ	-1.55
ACT vs <i>Planck</i> +SPT	18.4	0.52%	2.79σ	-6.22
ACT + SPT vs <i>Planck</i>	12.2	5.81%	1.90σ	-3.09
ACT+ <i>Planck</i> vs SPT	10.3	11.09%	1.59σ	-2.17

and by construction will have a one-dimensional marginalized tension of χ , which in the case of consistency takes the value $t \sim \sqrt{d} \pm 1/\sqrt{2}$. Maximum tension coordinates will be discussed in greater detail in an upcoming work [36].

Marginalized one and two-dimensional projections of the posteriors and the *Planck*-ACT tension coordinate are summarized in Fig. 1. It is easy for the eye to be drawn to certain projections where the marginalized tension is large, but as the maximum tension coordinate demonstrates, these can be misleading. We emphasize that the Suspiciousness synthesizes all of the posterior information correctly into a single intuitive summary statistic.

Comparing ACT with SPT [37], we find a slightly lower mild-to-moderate tension of 2.37σ ($p = 1.8\%$). Interestingly, comparing SPT with *Planck* we find no significant evidence for tension ($p = 16.8\%$), in contradiction with some of the historical literature [18], and in agreement with [32].

Since SPT and *Planck* are consistent, we may confidently combine these datasets. In the absence of a full pipeline run, we combine the Gaussian posterior approximations using Eqs. (14)–(20) from H19. This *Planck*+SPT combination is 2.79σ in tension with ACT ($p = 0.52\%$), well into the “moderate” regime.

Since ACT is in mild-to-moderate tension with both *Planck* and SPT, we should be suspicious of combining it with either, but when we do, as in the final two rows of Table I, we find no significant evidence for tension, although still higher than when comparing *Planck* and SPT.

TABLE II. We compare the tension computed using the full non-Gaussian expression from Eq. (4), and the tension computed via the Gaussian approximation. Note that in both cases, for this application since all parameters are well-constrained, all that is required are publicly available MCMC chains.

ACT vs <i>Planck</i> tension metric	p	Tension	$\log S$
True Suspiciousness Eq. (4)	0.57%	2.76σ	-6.10
Gaussian approximation Eq. (5)	0.86%	2.63σ	-5.60

In Table I we also report the χ^2 values for each data combination, and the Suspiciousness $\log S$ for reference. As $\log S$ can be regarded as the most conservative value $\log R$ can take by adjusting priors, it is interesting that all values are negative, reflecting the fact that all of the tension probabilities are a little low, when one would traditionally expect p to be uniformly distributed in a frequentist sense, and in general for $d = 6$ one would expect positive values of $\log S$ 58% of the time.

In Table II we compare the full non-Gaussian tension evaluated using Eq. (4) and find the Gaussian approximation to be a slight underestimate of the tension.

V. CONCLUSIONS

In general the causes of tension can be one of three things: (a) statistical fluctuation (b) systematics in at least one of the experiments (c) evidence for new physics. Given that we confidently launch manned space missions with higher failure rates than these tensions,³ as Bayesians we should be very concerned that our CMB measurements are in this much disagreement, so should view statistical fluctuations at this level as a very unsatisfactory explanation.

The general view (or hope) of many members of the cosmological community at the moment is that the cause of all of these tensions is likely a combination of (b) and (c), and before anyone can claim any kind of new physics we need to get a stronger handle on the systematics in many of our cosmological probes.

As mentioned earlier, this analysis can and will be improved by using a full pipeline of evidences and KL divergences computed using nested sampling [38], as well as using techniques that are specialized for dealing with correlated datasets. However, we would like to draw practitioners’ attention in particular to Eq. (4), which allows them to compute the Suspiciousness using only MCMC chains.

In this paper we do not seek to pass judgement on any of the *Planck*, ACT, or SPT analyses. Indeed, it could be argued that given the quality of all three analyses, it is more likely that these discrepancies indicate a problem with the underlying cosmology, rather than any of the independent pipelines. Combined with the many other tensions emerging between other datasets, the discrepancy quantified in this work lends credence to the possibility that before long we may yet see a paradigm shift in our understanding of the universe.

ACKNOWLEDGMENTS

W. H. thanks Gonville & Caius College for their support via a Research Fellowship. P. L. thanks STFC & University College London for their support via a STFC Consolidated Grant. Many thanks are accorded to Lukas Hergt for invaluable contributions to the ANESTHETIC package, and to Erminia Calabrese and Daan Meerburg for comments on an early draft.

³“Bet someone’s life” probabilities can be computed using data from <http://www.spacelaunchreport.com/logyear.html>

- [1] D. Scott, The Standard Model of cosmology: A skeptic's guide, *Proc. Int. Sch. Phys. Fermi* **200**, 133 (2020).
- [2] G. Efstathiou, H_0 revisited, *Mon. Not. R. Astron. Soc.* **440**, 1138 (2014).
- [3] E. Mörtzell and S. Dhawan, Does the Hubble constant tension call for new physics?, *J. Cosmol. Astropart. Phys.* **09** (2018) 025.
- [4] A. G. Riess, S. Casertano, W. Yuan, L. M. Macri, and D. Scolnic, Large magellanic cloud cepheid standards provide a 1% foundation for the determination of the Hubble constant and stronger evidence for physics beyond Λ CDM, *Astrophys. J.* **876**, 85 (2019).
- [5] M. Douspis, L. Salvati, and N. Aghanim, On the tension between large scale structures and cosmic microwave background, [arXiv:1901.05289](https://arxiv.org/abs/1901.05289).
- [6] H. Hildebrandt *et al.*, KiDS + VIKING-450: Cosmic shear tomography with optical and infrared data, *Astron. Astrophys.* **633**, A69 (2020).
- [7] W. Handley, Curvature tension: Evidence for a closed universe, *Phys. Rev. D* **103**, L041301 (2021).
- [8] E. Di Valentino, A. Melchiorri, and J. Silk, Planck evidence for a closed Universe and a possible crisis for cosmology, *Nat. Astron.* **4**, 196 (2020).
- [9] G. Efstathiou and S. Gratton, The evidence for a spatially flat Universe, *Mon. Not. R. Astron. Soc.* **496**, L91 (2020).
- [10] T. Charnock, R. A. Battye, and A. Moss, Planck data versus large scale structure: Methods to quantify discordance, *Phys. Rev. D* **95**, 123535 (2017).
- [11] W. Lin and M. Ishak, Remarks on measures of inconsistency, [arXiv:1909.10991](https://arxiv.org/abs/1909.10991).
- [12] W. Handley, Anesthetic: Nested sampling visualisation, *J. Open Source Softw.* **4**, 1414 (2019).
- [13] ACT Collaboration, The Atacama cosmology telescope: DR4 maps and cosmological parameters, *J. Cosmol. Astropart. Phys.* **12** (2020) 047.
- [14] ACT Collaboration, The Atacama cosmology telescope: A measurement of the cosmic microwave background power spectra at 98 and 150 GHz, *J. Cosmol. Astropart. Phys.* **12** (2020) 045.
- [15] ACT Collaboration, The Atacama cosmology telescope: Arcminute-resolution maps of 18,000 square degrees of the microwave sky from ACT 2008-2018 data combined with Planck, *J. Cosmol. Astropart. Phys.* **12** (2020) 046.
- [16] Planck Collaboration, Planck 2018 results. V. CMB power spectra and likelihoods, *Astron. Astrophys.* **641**, A5 (2020).
- [17] Planck Collaboration, Planck 2018 results. VI. Cosmological parameters, *Astron. Astrophys.* **641**, A6 (2020).
- [18] SPT Collaboration, Measurements of the temperature and E-mode polarization of the CMB from 500 square degrees of SPTpol data, *Astrophys. J.* **852**, 97 (2018).
- [19] W. Handley and P. Lemos, Quantifying tensions in cosmological parameters: Interpreting the DES evidence ratio, *Phys. Rev. D* **100**, 043504 (2019).
- [20] P. Marshall, N. Rajguru, and A. Slosar, Bayesian evidence as a tool for comparing datasets, *Phys. Rev. D* **73**, 067302 (2006).
- [21] H. Jeffreys, *Theory of Probability*, International series of monographs on physics (Clarendon Press, 1939), <https://books.google.co.uk/books?id=Bu1rAAAAIAAJ>.
- [22] DES Collaboration, Dark energy survey year 1 results: Cosmological constraints from galaxy clustering and weak lensing, *Phys. Rev. D* **98**, 043526 (2018).
- [23] S. Kullback and R. A. Leibler, On information and sufficiency, *Ann. Math. Stat.* **22**, 79 (1951).
- [24] P. Lemos, F. Köhlinger, W. Handley, B. Joachimi, L. Whiteway, and O. Lahav, Quantifying suspiciousness within correlated data sets, *Mon. Not. R. Astron. Soc.* (2020).
- [25] C. Heymans *et al.*, KiDS-1000 cosmology: Multi-probe weak gravitational lensing and spectroscopic galaxy clustering constraints, *Astron. Astrophys.* **646**, A140 (2021).
- [26] L. T. Hergt, W. J. Handley, M. P. Hobson, and A. N. Lasenby, Bayesian evidence for the tensor-to-scalar ratio r and neutrino masses m_ν : Effects of uniform vs logarithmic priors, [arXiv:2102.11511](https://arxiv.org/abs/2102.11511).
- [27] M. Raveri, G. Zacharegkas, and W. Hu, Quantifying concordance of correlated cosmological data sets, *Phys. Rev. D* **101**, 103527 (2020).
- [28] M. Raveri and W. Hu, Concordance and discordance in cosmology, *Phys. Rev. D* **99**, 043506 (2019).
- [29] Y. Park and E. Rozo, Concordance cosmology?, *Mon. Not. R. Astron. Soc.* **499**, 4638 (2020).
- [30] F. Köhlinger, B. Joachimi, M. Asgari, M. Viola, S. Joudaki, and T. Tröster, A Bayesian quantification of consistency in correlated data sets, *Mon. Not. R. Astron. Soc.* **484**, 3126 (2019).
- [31] W. Handley and P. Lemos, Quantifying dimensionality: Bayesian cosmological model complexities, *Phys. Rev. D* **100**, 023512 (2019).
- [32] Planck Collaboration, Planck 2018 results. I. Overview and the cosmological legacy of Planck, *Astron. Astrophys.* **641**, A1 (2020).
- [33] E. Calabrese, A. Slosar, A. Melchiorri, G. F. Smoot, and O. Zahn, Cosmic microwave weak lensing data as a test for the dark universe, *Phys. Rev. D* **77**, 123531 (2008).
- [34] G. Efstathiou and S. Gratton, A detailed description of the CamSpec likelihood pipeline and a reanalysis of the Planck high frequency maps, [arXiv:1910.00483](https://arxiv.org/abs/1910.00483).
- [35] Planck Collaboration, Planck 2018 results. VIII. Gravitational lensing, *Astron. Astrophys.* **641**, A8 (2020).
- [36] W. Handley, L. Lau, P. Lemos, and W. Barker, Maximum tension coordinates with cosmological applications (to be published).
- [37] E. Di Valentino, S. Galli, M. Lattanzi, A. Melchiorri, P. Natoli, L. Pagano, and N. Said, Tickling the CMB damping tail: Scrutinizing the tension between the atacama cosmology telescope and south pole telescope experiments, *Phys. Rev. D* **88**, 023501 (2013).
- [38] J. Skilling, Nested sampling for general Bayesian computation, *Bayesian Anal.* **1**, 833 (2006).