Blinding multiprobe cosmological experiments

J. Muir, G.M. Bernstein, D. Huterer, F. Elsner, E. Krause, A. Roodman, S. Allam, J. Annis, S. Avila, K. Bechtol, et al.

Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **494,** 4454–4470 (2020)
Advance Access publication 2020 April 15

doi:10.1093/mnras/staa965

Blinding multiprobe cosmological experiments

J. Muir[1,2★] G. M. Bernstein,[3] D. Huterer,[2] F. Elsner[4,5] E. Krause,[6] A. Roodman,[1,7] S. Allam,[8] J. Annis,[8] S. Avila,[9] K. Bechtol,[10,11] E. Bertin,[12,13] D. Brooks,[4] E. Buckley-Geer,[8] D. L. Burke,[1,7] A. Carnero Rosell,[14,15] M. Carrasco Kind,[16,17] J. Carretero,[18] R. Cawthon,[11] M. Costanzi,[19,20] L. N. da Costa,[15,21] J. De Vicente,[14] S. Desai,[22] J. P. Dietrich,[23,24] P. Doel,[4] T. F. Eifler,[6,25] S. Everett,[26] P. Fosalba,[27,28] J. Frieman,[8,29] J. García-Bellido,[9] D. W. Gerdes,[30,2] D. Gruen,[31,1,7] R. A. Gruendl,[16,17] J. Gschwend,[15,21] W. G. Hartley,[4,32] D. L. Hollowood,[26] D. J. James,[33] M. Jarvis,[3] K. Kuehn,[34,35] N. Kuropatkin,[8] O. Lahav,[4] M. March,[3] J. L. Marshall,[36] P. Melchior,[37] F. Menanteau,[16,17] R. Miquel,[18,38] R. L. C. Ogando,[15,21] A. Palmese,[8,29] F. Paz-Chinchón,[16,17] A. A. Plazas,[37] A. K. Romer,[39] E. Sanchez,[14] V. Scarpine,[8] M. Schubnell,[2] S. Serrano,[27,28] I. Sevilla-Noarbe,[14] M. Smith,[40] E. Suchyta,[41] G. Tarle,[2] D. Thomas,[42] M. A. Troxel,[43] A. R. Walker,[44] J. Weller,[23,45,46] W. Wester,[8] J. Zuntz[47] and (DES Collaboration)

Affiliations are listed at the end of the paper

Accepted 2020 April 1. Received 2020 April 1; in original form 2019 November 15

ABSTRACT

The goal of blinding is to hide an experiment's critical results – here the inferred cosmological parameters – until all decisions affecting its analysis have been finalized. This is especially important in the current era of precision cosmology, when the results of any new experiment are closely scrutinized for consistency or tension with previous results. In analyses that combine multiple observational probes, like the combination of galaxy clustering and weak lensing in the Dark Energy Survey (DES), it is challenging to blind the results while retaining the ability to check for (in)consistency between different parts of the data. We propose a simple new blinding transformation, which works by modifying the summary statistics that are input to parameter estimation, such as two-point correlation functions. The transformation shifts the measured statistics to new values that are consistent with (blindly) shifted cosmological parameters while preserving internal (in)consistency. We apply the blinding transformation to simulated data for the projected DES Year 3 galaxy clustering and weak lensing analysis, demonstrating that practical blinding is achieved without significant perturbation of internal-consistency checks, as measured here by degradation of the χ^2 between the data and best-fitting model. Our blinding method's performance is expected to improve as experiments evolve to higher precision and accuracy.

Key words: methods: data analysis – methods: numerical – methods: statistical – cosmology: observations; large-scale structure of Universe.

1 INTRODUCTION

The practice of blinding against human bias in data analysis is standard in many areas of science. The goal is to prevent the scientists from biasing their analysis toward results that are theoretically expected or, more generally, deemed to be likely or correct. In experimental particle physics, strategies for blinding are manyfold and have been honed since their earliest application decades ago (Arisaka et al. 1993). Blinding strategies in particle physics include hiding the signal region, offsetting parameters in the analysis by a hidden constant, and adding or removing

★ E-mail: jlmuir@stanford.edu

events from the analysis (for a review, see Klein & Roodman 2005).

Blinding started to be applied to astrophysics and cosmology only relatively recently. The first application to cosmology was described in Conley et al. (2006), which reports on an analysis of magnitude-redshift data of Type Ia supernovae (SNe Ia). In that study, the full analysis was performed with unknown offsets added to the key cosmological parameters, $\Omega_{\rm M}$ and Ω_Λ, until unblinding revealed final parameter values. Many SN Ia analyses have adopted some variation of this blinding approach since (e.g. Kowalski et al. 2008; Suzuki et al. 2012; Betoule et al. 2014; Rubin et al. 2015; Zhang et al. 2017; Abbott et al. 2019). More recently, blinding has been regularly applied to analyses involving strong gravitational lensing (Suyu et al. 2013, 2017), as well as cosmological inferences from weak gravitational lensing observations (e.g. Heymans et al. 2012; von der Linden et al. 2014; Kuijken et al. 2015; Blake et al. 2016; Hildebrandt et al. 2017; Troxel et al. 2018; Hamana et al. 2020).

For cosmological analyses in general, direct application of blinding techniques from experimental particle physics is not feasible due to numerous differences between cosmological observations and particle-physics experiments. First, there is no clear division of the data space into a 'signal' region that can be hidden versus a 'control' region that can be used for all validation tests. A second significant challenge arises from the fact that many cosmological inferences are now produced by combination of multiple 'probes', i.e. summary statistics of diverse forms of measurement of different classes of objects in the sky. For example, Abbott et al. (2018) present a combined analysis of an observable vector $\boldsymbol{d}$ composed of three two-point correlation functions (2PCFs) measured from the first year of Dark Energy Survey (DES) data: the angular correlation of galaxy positions via $w(\theta)$, the angular correlation of weak lensing shears via $\xi_\pm(\theta)$, and the cross-correlation between galaxy positions and shears via $\gamma_t(\theta)$. There is much degeneracy in how these diverse measurements contain cosmological information, which means that a simple blinding operation applied to one measured quantity can transform valid data into blinded data that are readily recognized as inconsistent with any viable cosmology.

The simplest form of blinding, which was used in the DES Year 1 galaxy clustering and weak lensing analysis, is to hide from users the values of the cosmological parameters that arise from the final inference, e.g. by shifting all values in any human-readable results. The risk of accidental revelation of the true parameter estimates is high, however, if the blinding code is mistakenly omitted. The temptation for experimenters to peek at the true results is also high when the 'curtain' is so thin. Furthermore, in this scenario, the blinding can potentially be compromised if anyone plots a theoretical model on top of the measured summary statistics. It is therefore an advantage to apply a blinding transformation at an earlier stage, when more steps are required to produce unblinded results in a form that an experimenter can recognize as conforming to their biases or not.

In this paper, we propose a method for blinding multiprobe cosmological analyses by altering the summary statistics which are used as input for parameter estimation. The technique is very simply described by equation (7). This technique has the advantages of being applicable to observable vectors of arbitrary complexity while preserving internal consistency checks, and also of insuring that the inference code never even produces the true cosmological parameters until the collaboration agrees to unblind. We are specifically developing and testing the performance of this blinding scheme for the DES Year 3 combined probe analysis, but the ideas we present could, in principle, be applied to any cosmological analysis. Accordingly, we frame our discussion in terms of a generic experiment, beginning in Section 2 with a discussion of general considerations for blinding and how to assess whether a blinding scheme can be successful. This is followed by Section 3 where we introduce our summary-statistic-based blinding transformation. Then, Section 4 describes the transformation's application within the DES analysis pipeline as well as the results of our tests of its performance for a simulated DES Year 3 galaxy clustering and weak lensing analysis. We conclude in Section 5. The data associated with the tests described below are available upon request.

2 PRIOR AND PREJUDICE: CONSIDERATIONS FOR BLINDING

Broadly speaking, the goal of blinding is to change or hide the output of an analysis in a way that still allows experimenters to perform validation checks on the analysis pipeline and data. Thus, in order to be effective, a given blinding scheme must fulfill these requirements: it must be capable of altering the analysis' output enough to overcome biases, and, additionally, must preserve the properties of the data that are to be used in validation tests. Below we present three criteria that can guide the determination of whether a given transformation of data can successfully blind an analysis.

2.1 Criterion I: concealing the true results

Let us assume that the experiment produces a vector $\hat{\boldsymbol{d}}$ of observed quantities, and we wish to constrain the parameters Θ of a model $\boldsymbol{d}(\Theta)$ for these data. The parameters can include astrophysical and instrumental nuisance parameters as well as the cosmological parameters of interest. There will always be some prior probability, $\pi(\Theta)$, that expresses the physical bounds of our model (e.g. $\Omega_{\rm m} > 0$) and results of trusted previous experimentation. In a Bayesian view, the purpose of the experiment is to produce a likelihood function $\mathcal{L}(\hat{\boldsymbol{d}}|\Theta)$, which is combined with the prior to produce a posterior measure of belief across the model space $\mathcal{M}$ spanned by parameter vector $\boldsymbol{\Theta}$, $P(\Theta|d) \propto \mathcal{L}(\hat{\boldsymbol{d}}|\Theta)\,\pi(\Theta)$. One easily visualized variant of the prior is to have $\mathcal{M}$ be uniform over some range of parameter values and zero elsewhere, i.e. $\mathcal{M}$ is the parameter space encompassing all parameters Θ considered feasible.

The experimenters may additionally harbour prejudices about the 'correct' values of the parameters; for instance, that they should agree with some theoretical framework such as a flat Universe, or that they should agree with some previous experiment that one is trying to confirm. We can express these prejudices with another (albeit, difficult to quantify) probability function Prej(Θ). It could for example be a uniform distribution over some region $\mathcal{M}_{\rm Prej} \subset \mathcal{M}$. Note that in this framing, one must make a decision regarding previous experiments' results: either we accept them as true and place them in π; or we are using their comparison to our results as a test of our model, in which case we must be wary of confirmation bias and should place them in Prej.

The danger of experimenter bias arises when choices about the analysis process are made, consciously or otherwise, on the basis of whether the experiment's results conform to the prejudices, i.e. whether $\Theta \in \mathcal{M}_{\rm Prej}$. To confound the experimenter bias, a blinding procedure will apply a transformation

$$\hat{\boldsymbol{d}} \rightarrow \hat{\boldsymbol{d}}_{\rm bl} = B(\hat{\boldsymbol{d}}) \tag{1}$$

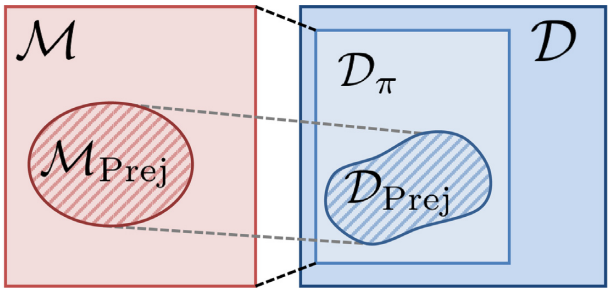

Figure 1. Cartoon of model and data spaces that we consider when thinking about how to blind an analysis, as discussed in Section 2.1. $\mathcal{M}$ is the space of all viable model parameter sets Θ, which projects on to the observable-vector space $\mathcal{D}_\pi \subset \mathcal{D}$, where $\mathcal{D}$ is the space of all possible observable vectors. $\mathcal{M}_{\rm Prej}$ is a region in parameter space associated with what we refer to as the 'prejudice' distribution, describing experimenters' preconceived expectations for where Θ is likely to be. This subset of parameter space projects on to $\mathcal{D}_{\rm Prej} \subset \mathcal{D}_\pi$. An effective blinding transformation must have the possibility of moving the observable vector $\hat{\boldsymbol{d}}$ in or out of $\mathcal{D}_{\rm Prej}$ without moving it out of the prior space $\mathcal{D}_\pi$.

to the data before the experimenters perform analyses. The first critical property of B is therefore that it must confound the experimenter's ability to know whether the data are consistent or inconsistent with their prejudices. For example, if we take the maximum-posterior parameter values for blinded and unblinded data

$$\Theta_{\rm unbl} = \underset{\Theta}{\rm argmax} \left\{ P\left(\Theta|\hat{\boldsymbol{d}}\right)\right\},$$
$$\Theta_{\rm bl} = \underset{\Theta}{\rm argmax} \left\{ P\left(\Theta|B(\hat{\boldsymbol{d}})\right)\right\}, \quad (2)$$

then there must be a non-negligible chance that either

$$\mathrm{Prej}(\Theta_{\rm unbl}) \gg \mathrm{Prej}(\Theta_{\rm bl}) \quad \text{or} \quad \mathrm{Prej}(\Theta_{\rm unbl}) \ll \mathrm{Prej}(\Theta_{\rm bl}). \quad (3)$$

A graphical illustration is given in Fig. 1: If we define $\mathcal{D}_{\rm Prej}$ as the region of data space $\mathcal{D}$ produced by parameter values within the prejudice region $\mathcal{M}_{\rm Prej}$, then the blinding transformation must be able to move data into and out of this region. The experimenters should believe that this is possible, but not know for certain whether it has happened. Because we are dealing with human psychology and prejudices which may not be quantifiable, we usually cannot create a strict numerical requirement to satisfy this criterion.

2.2 Criterion II: preserving the ability to check for errors

In addition to obscuring the true parameter output of an analysis, an effective blinding scheme must still allow experimenters to examine the data $\hat{\boldsymbol{d}}$, before unblinding, to uncover errors in their analysis procedure. A validation test is one whose failure indicates that data could not have been produced by any allowed parameters $\Theta \in \mathcal{M}$. A blinding transformation B should not alter the conclusions of validation tests.

There are a number of ways of stating this requirement. Sometimes the validation tests are expressed as some projection of the data on to a 'null test' $T(\hat{\boldsymbol{d}})$ such that

$$T(\hat{\boldsymbol{d}}) = 0 \quad \forall \Theta \in \mathcal{M}. \quad (4)$$

Many kinds of validation tests fall into this paradigm. For example, if T projects on to the B mode (divergence-free component) of weak lensing, it should be zero within errors. Another example is that a properly extinction-corrected galaxy survey should exhibit no statistically significant correlation between galaxy positions and star positions, so T in this case would be the angular star–galaxy correlation function. Or, T can measure the difference between observable vectors split by some property presumed to be uncorrelated with extra-galactic signals, such as seeing or the season when the data were collected. Another very generic test is to run the parameter inference on two subsets of the observable vector and check that the results are consistent with common Θ. Allowances must of course be made for the expected noise in the null test output at fixed Θ. Generally speaking, a useful blinding transformation must yield

$$T(\hat{\boldsymbol{d}}_{\rm bl}) \sim 0 \quad \text{if and only if} \quad T(\hat{\boldsymbol{d}}) \sim 0 \quad (5)$$

where the $\sim$ sign implies consistency with zero within measurement errors.

More generally, $B(\hat{\boldsymbol{d}})$ should map the allowed region $\mathcal{D}_\pi$ on to itself, and, likewise for its complement, the disallowed region $\tilde{\mathcal{D}}_\pi$. Equivalently, the maximum-posterior values from equations (2) should obey

$$\frac{\mathcal{L}\left(\hat{\boldsymbol{d}}_{\rm bl}|\Theta_{\rm bl}\right)}{\mathcal{L}\left(\hat{\boldsymbol{d}}_{\rm unbl}|\Theta_{\rm unbl}\right)} \approx 1. \quad (6)$$

In other words, blinding transformation should not significantly change the maximum likelihood in the parameter space. A transformation satisfying this requirement will ensure that blinding will not alter experimenters' judgment about whether there are flaws in the data.

2.2.1 *Model dependence of validation tests*

It is important to note that defining validation tests requires one to make implicit modeling choices, and defining a blinding procedure that preserves the result of those tests can only produce shifts in parameter space that respects those choices. When constructing an analysis pipeline, it is therefore important to carefully consider what measurements will be considered results and which can be used as checks on the performance of the analysis pipeline. In other words, defining validation criteria and a blinding scheme that preserves them requires one to specify the space of models that are considered viable.

The only case in which a purely internal validation comparison can be made without reference to a model is if the exact same observable is measured twice [e.g. the amplitude of a particular cosmic microwave background (CMB) harmonic measured at a particular frequency]. Any time two distinct quantities enter the observable vector, a model is required to constrain their joint distribution. For example, suppose we compare the high- and low-redshift halves of a supernova *Hubble* diagram. If both halves are fit with a Lamba cold dark matter (ΛCDM) model and the data truly are from a ΛCDM universe, then analysing the two halves separately should produce consistent cosmology results, making this comparison a useful validation test. If, however, the universe is not described by ΛCDM, then the high/low-z split can yield inconsistent results even in the absence of processing errors. (Note that the original discovery of dark energy was effectively a demonstration that fitting supernovae data with $\Lambda = 0$ produced this kind of mismatch.)

More subtly, a closer examination of the B-mode null test described above (below equation 4) is another demonstration of how even nominally 'simple' tests can be model-dependent. Though it

is true that at leading-order weak lensing in GR should have no divergence-free component, B modes can, in fact, be created by galaxy intrinsic alignments (IA) or modifications to GR. Thus, the test requiring measured B modes to be zero within errors can be more accurately rephrased as the requirement that the measured B modes be consistent with allowed IA and gravity models.

2.3 Criterion III: feasible implementation

Before we can begin determining the effectiveness of a blinding transformation B, we must first choose what data will be transformed. As a concrete example, for imaging surveys like DES the data start as pixel values, which are converted into catalogued galaxy fluxes, shapes, etc. Those catalogues are, in turn, converted into summary statistics such as the photometric redshift distributions $n(z)$ and tomographic weak lensing correlation functions $\xi_{\pm}(\theta)$. The summary statistics are then finally used to obtain the parameter estimates $\hat{\Theta}$ themselves.

The simplest case is simply for B to operate on the $\hat{\Theta}$. As noted in Section 1, this 'parameter shift' method has been used frequently, but it has the drawback of being a very thin cover over the truth. It fails, for example, if the experimenters become familiar with the relation between the observables and the cosmological parameters and are viewing how changes to the analysis impact the observables. It can also be difficult to employ this method while insuring that multiple probes are consistent with a common model. There is incentive to move the blinding transformation to an earlier stage of the analysis, where the scientists are less likely to be able to recognize whether their prejudices have been confirmed.

Blinding by alteration of the pixel data is probably impossible, apart from substituting an entire set of simulated data for the real one. Blinding at the catalogue level is possible in some cases, namely when a change in a cosmological parameter maps directly into a change in some catalogued galaxy property. For the DES Year 1 shear-only analyses (Troxel et al. 2018), all galaxy ellipticities (and hence all inferred weak lensing shears) were scaled by an unknown multiplicative factor. This is approximately equivalent to a rescaling of σ_8, though only in the linear regime. The possibility of catalogue-level transformations becomes remote, however, as we conduct multiprobe experiments with many correlated summary statistics, and as multiple model parameters require blinding. We have not been able to find a catalogue transformation that preserves the validity of the data for the DES combined galaxy clustering and weak lensing analysis. This has motivated us to develop an approach to blinding, which relies on a transformation of the summary statistics, described in more detail below.

Recently, Sellentin (2020) proposed a likelihood-level blinding via modifications of the full covariance matrix of the observable vector. It will be of interest to see if this alternative blinding method robustly satisfies criterion II from above – preserving the results of all validation tests – given that Sellentin's (2020) alteration to the covariance matrix is dependent on the observed data.

3 PROPOSED METHOD: BLINDING BY MODIFYING SUMMARY STATISTICS

Here we propose a method for consistently blinding cosmological analyses by transforming the summary statistics used as input for parameter estimation. Because parameter estimation is done by comparing measured summary statistics to theoretical predictions, the software infrastructure for an experiment will naturally include tools for computing model predictions at various points in parameter space. Our blinding transformation makes use of those tools to translate shifts in parameter space to changes in the summary statistics.

The blinding transformation works as follows. Let $\hat{d}_i$ be element i of a measured observable vector, and let $d_i(\Theta)$ be the theoretically computed (noiseless) value of that same element for model parameters Θ. We choose a known reference model $\Theta_{\rm ref}$ and a blinding shift $\Delta\Theta$ in the cosmological parameters. The blinding operation is then a simple modification of each element $\hat{d}_i$ of the observable vector,

$$B(\hat{d}_i) = \hat{d}_i + f_i^{(\rm add)},$$
$$f_i^{(\rm add)} = d_i\left(\Theta_{\rm ref} + \Delta\Theta\right) - d_i\left(\Theta_{\rm ref}\right). \quad (7)$$

If the expected noise level on $\hat{\boldsymbol{d}}$ does not vary much across the parameter shift $\Delta\Theta$, then it is true by construction that B will map data generated at $\Theta_{\rm obs}$ into viable data for $\boldsymbol{\Theta}_{\rm bl} = \boldsymbol{\Theta}_{\rm obs} + \Delta\Theta$ if the truth ($\Theta_{\rm obs}$) is sufficiently close to the reference cosmology ($\Theta_{\rm ref}$). However, because $\Theta_{\rm obs}$ is not known (in fact, this whole exercise is designed to obscure it!) and because the observable vector generally is not actually linear in parameters, it is not guaranteed that this blinding transformation will satisfy the necessary criteria for successful blinding. Its application to a given observable vector and parameter space thus requires numerical validation.[1]

3.1 Procedure for blinding at the level of summary statistics

An overview of the procedure for summary-statistic blinding is as follows:

(i) Choose a reference cosmology (and nuisance parameters) $\Theta_{\rm ref}$ in the middle of the range of models considered feasible truths.

(ii) Select a (blind) shift $\Delta\Theta$ from a distribution broader than the preconceptions causing the confirmation bias. For example, if there is a theoretical prejudice that the dark energy equation of state parameter is $w = -1$, then Δw should be capable of shifts four to five times the experiment's forecasted uncertainty in w.

(iii) For each summary statistic d_i being used for cosmological inference, calculate the blinding factor f_i using equation (7).

(iv) Hide the real data $\hat{d}_i$ and give experimenters the shifted values $B(\hat{d}_i)$ as per equation (7) with which to conduct all validation tests.

(v) After passing validation tests, unblind by using the original unblinded data $\hat{\boldsymbol{d}}$ to repeat the inference of Θ.

3.2 Evaluating performance

The blinding technique that we propose is fully described by equation (7). In practice, the implementation of our blinding algorithm depends on the choice of the summary statistic to which the blinding factors $\boldsymbol{f}^{(\rm add)}$ are applied, the reference parameters $\Theta_{\rm ref}$, as well as the probability distribution from which parameter shifts $\Delta\Theta$ are drawn. In order to test the performance for a given set of these choices, we must show the following:

[1]For some observables, it might be possible to blind using a multiplicative transformation, multiplying the observable vector entries i by $f_i^{(\rm mult)} = d_i\left(\Theta_{\rm ref} + \Delta\Theta\right)/d_i\left(\Theta_{\rm ref}\right)$. Our tests show, however, that this would rescale the noise in the observable vector as well as the signal, and would lead to unpredictable behaviour if any components d_i are close to zero. Thus, in most cases equation (7)'s additive transformation will be preferable.

(1) the blinding transformation is able generate shifts in best-fitting model parameters large enough to overcome experimenters' potential biases, as described in Section 2.1, and

(2) the blinded observable vector $\hat{\boldsymbol{d}}_{\rm bl}$ is consistent with data that could be produced by *some* set of allowed model parameters, as discussed in Section 2.2.

We can test both of these requirements by analysing simulated data according to the procedure below.

(i) We choose a reasonable reference cosmology $\Theta_{\rm ref}$ as well as an ensemble of 'true' parameters associated with observed unblinded data $\{\Theta^{(a)}_{\rm obs}\}$, where a labels the realization. For these realizations, we also select an ensemble blinding shifts $\{\Delta\Theta^{(a)}\}$. For example, we may choose to use the reference cosmology with the dark energy equation of state $w_{\rm ref} = -1$, and in some realizations of our blinding test, we could choose $w^{(a)}_{\rm obs} = -0.875$ and $\Delta w^{(a)} = -0.031$.

(ii) For each realization, we generate a noiseless synthetic measured observable vector by computing the theory prediction for the data at the input cosmology:

$$\hat{\boldsymbol{d}}^{(a)} = \boldsymbol{d}\left(\Theta^{(a)}_{\rm obs}\right). \tag{8}$$

In our representative example above, this corresponds to, for example, a predicted weak lensing shear 2PCF evaluated at $w^{(a)}_{\rm obs} = -0.875$.

(iii) We then blind that data vector using $\Theta_{\rm ref}$ and $\Delta\Theta^{(a)}$ according to the transformation equation (7) to obtain $\hat{\boldsymbol{d}}^{(a)}_{\rm bl}$. That is, we evaluate

$$\hat{\boldsymbol{d}}^{(a)}_{\rm bl} = \boldsymbol{d}\left(\Theta^{(a)}_{\rm obs}\right) + \boldsymbol{d}(\Theta_{\rm ref} + \Delta\Theta^{(a)}) - \boldsymbol{d}(\Theta_{\rm ref}). \tag{9}$$

In our example, this corresponds to the sum of the 2PCF for $w^{(a)}_{\rm obs} = -0.875$ and one for $w_{\rm ref} + \Delta w^{(a)} = -0.906$, minus the 2PCF for $w_{\rm ref} = -1$.

(iv) By performing parameter estimation on $\hat{\boldsymbol{d}}^{(a)}$, we can find associated unblinded best-fitting parameters $\Theta^{(a)}_{\rm unbl}$. (For noiseless data, we expect $\Theta^{(a)}_{\rm unbl} = \Theta^{(a)}_{\rm obs}$.) Likewise, we can find the blinded best-fitting parameters $\Theta^{(a)}_{\rm bl}$ by performing parameter estimation on $\hat{\boldsymbol{d}}^{(a)}_{\rm bl}$.

Studying the distribution of these best-fitting parameters for such a simulated analyses allows us to assess the performance of the blinding transformation. Point (i) from above (that blinding must be able to produce large enough shifts in parameter space) is straightforward to check. Generally, we expect that the input blinding shift will determine the shift in output best-fitting parameters:

$$\Theta_{\rm bl} - \Theta_{\rm unbl} \approx \Delta\Theta. \tag{10}$$

If this is true, we can ensure the blinding transformation satisfies this requirement simply by drawing $\Delta\Theta$ from a wide enough probability distribution in $\mathcal{M}$. By analysing an ensemble of simulated observable vectors, we can explicitly check the extent to which equation (10) holds. It is worth noting here that it does not matter if the relation in equation (10) strictly holds: Blinding can still be effective as long as the output shifts $\Theta_{\rm bl} - \Theta_{\rm unbl}$ and input shifts $\Delta\Theta$ span a comparable region of parameter space.

For point (2), we propose using the quantity $\Delta\chi^2$, defined below, as a metric for testing whether the blinding transformation defined in equation (7) preserves the results of validation tests.[2] This statistic is defined as the difference between the minimum χ^2 for the model's fit to blinded data and that of the fit to unblinded data:

$$\Delta\chi^2 \equiv \left\{-2\ln\mathcal{L}\left(B(\hat{\boldsymbol{d}})|\Theta_{\rm bl}\right)\right\} - \left\{-2\ln\mathcal{L}\left(\hat{\boldsymbol{d}}|\Theta_{\rm unbl}\right)\right\}. \tag{11}$$

It is a measure of the extent to which blinding preserves the internal consistency of the different components of the observable vector. In other words, it quantifies how much of the error in the model's fit to blinded data comes from the blinding procedure itself. If we can confirm that $\Delta\chi^2$ is sufficiently small for all realizations in our ensemble of simulated observable vectors, we can ensure that the blinding transformation satisfies equation (6) for the set of input parameters and blinding shifts considered.

3.3 Leading-order performance

In the context of the evaluation metric described above, a perfect blinding technique can shift the inferred parameters by the bias-defeating amount with $\Delta\chi^2 = 0$, i.e. no change in the degree to which data obey the model. It is clear that equation (7) will be perfect if the data depend on the model in a purely linear fashion. Furthermore, the parameter shift will be simple, i.e. equation (10) will attain equality.

Indeed, in this linear regime with fixed covariance, blinding via equation (7) guarantees the more general statement that

$$\chi^2\left[B(\hat{\boldsymbol{d}})|\Theta_{\rm obs} + \Delta\Theta\right] = \chi^2(\hat{\boldsymbol{d}}|\Theta_{\rm obs}), \tag{12}$$

which, in turn, will lead to the Bayesian evidence

$$p(\hat{\boldsymbol{d}}) = \int \mathrm{d}\Theta\, \mathcal{L}(\hat{\boldsymbol{d}}|\Theta)p(\Theta) \tag{13}$$

being invariant under the blinding transformation to the extent that the prior $p(\Theta)$ is invariant under shift by $\Delta\Theta$. In this limit of fixed multivariate Gaussian noise and linear parameter shifts, the additive blinding yields data that are fully indistinguishable from a shift in the truth cosmology.

The blinding technique of Sellentin (2020) differs in that the measured observable vector $\hat{\boldsymbol{d}}$ is left unchanged, but the covariance matrix undergoes a data-dependent transformation $C_d \to \check{C}_d(\hat{\mathbf{d}})$ such that the blinded $\check{\chi}^2(\hat{\boldsymbol{d}}|\Theta)$ is guaranteed to satisfy

$$\check{\chi}^2(\hat{\boldsymbol{d}}|\Theta_{\rm ref} + \Delta\Theta) = \chi^2(\hat{\boldsymbol{d}}|\Theta_{\rm ref}). \tag{14}$$

This is not quite the same condition as demanding $\Delta\chi^2 = 0$ in equation (11), since the latter operates at the best-fitting Θ in both the blinded and unblinded cases. The basic construction for $\check{C}_d$ does not guarantee that the analogous property to equation (12) will hold, i.e. we can have

$$\check{\chi}^2(\hat{\boldsymbol{d}}|\Theta + \Delta\Theta) \neq \chi^2(\hat{\boldsymbol{d}}|\Theta) \quad \text{if} \quad \Theta \neq \Theta_{\rm ref}, \tag{15}$$

even in the linear regime. We can expect that Sellentin's (2020) blinding is functionally perfect by the definition of $\Delta\chi^2$ in equation (11) being small, for the true Θ sufficiently close to $\Theta_{\rm ref}$. Numerical investigations will be of interest to see whether covariance-matrix blinding yields sufficiently small $\Delta\chi^2$ over the full range of Θ allowed by any particular experiment.

Moving beyond the linear regime, in Appendix A, we calculate the parameter shifts and $\Delta\chi^2$ induced by a quadratic dependence of data on parameters. The result is that the shift in best-fitting

[2]Of course, if one is considering applying this blinding procedure to an analysis which will use specific null or consistency tests as unblinding criteria, one should additionally check that the results of those tests are preserved.

parameters is no longer equal to $\Delta\Theta$, but rather acquires a leading term that is linear in the product $\Delta\Theta \times (\Theta_{\rm obs} - \Theta_{\rm ref})$. In other words, we should not expect equality in equation (10). Keep in mind though, as we noted above, fulfilling this equality is not a goal of blinding.

The more important result is the scaling

$$\Delta\chi^2 \sim |\Delta\Theta|^2\, |\Theta_{\rm obs} - \Theta_{\rm ref}|^2 \,/\, |C|\,, \tag{16}$$

where C is the covariance describing measurement errors on the parameters.

We can therefore expect that the blinding will succeed (by having insignificant $\Delta\chi^2$) within some sufficiently small region around the origin in the plane of blinding shift $\Delta\Theta$ and 'truth shift' $\Theta_{\rm obs} - \Theta_{\rm ref}$. This region must be large enough to allow the blinding shift to span $\mathcal{M}_{\rm Prej}$ and for the truth shift to span $\mathcal{M}$. Below, we will test whether this condition is met for the DES Year 3 analysis.

An important consequence of the scaling relations is that this blinding transformation will improve as our knowledge and experimental precision evolve. Let us assume that future experiments reduce measurement errors by a factor $\alpha < 1$ ($C \to \alpha^2 C$). To the extent that statistical rather than systematic uncertainties limit constraining power, the blinding shift necessary to defeat prejudices will shrink in concert ($\Delta\Theta \to \alpha\Delta\Theta$) as will the width of the priors to future experiments ($\Theta_{\rm obs} - \Theta_{\rm ref} \to \alpha(\Theta_{\rm obs} - \Theta_{\rm ref})$). Under equation (16), we see that $\Delta\chi^2 \propto \alpha^2$ under this evolution. Hence, to the extent that these trends hold, future experiments will, in fact, become easier to blind, given the same models and observables.[3]

4 APPLICATION: DARK ENERGY SURVEY ANALYSIS OF GALAXY LENSING AND CLUSTERING

Using the discussions in Sections 2 and 3 as a guide, we will now test the summary-statistic blinding transformation for the DES Year 3 galaxy clustering and weak lensing combined analysis. The goal of this exercise is twofold. First, it will serve as a concrete demonstration of how summary-statistic blinding can be implemented, and, secondly, we will validate the transformation's performance for use in the DES Year 3 analysis.

DES is an imaging survey that, over the course of six years, has measured galaxy positions and shapes in a 5000 deg^2 footprint in the southern sky. It is designed to use multiple observable probes to study the properties of dark energy and to otherwise test the standard cosmological model, ΛCDM. Those probes include galaxy clustering, weak lensing, supernovae, and galaxy clusters. Though the blinding transformation presented in this paper could be potentially useful for all of these cosmological observables, our focus in this paper is on the combined analysis of galaxy clustering and weak lensing shear.

[3]This scaling may not hold if, for example, tensions between previous experiments provoke prejudices which are large compared to the projected errors of the experiment. This would motivate blinding and truth shifts that shrink more slowly than measurement errors. If this shrinking is slow enough, preserving χ^2 could actually get more difficult. On the other hand, this would mean that we are in the fortunate position of having an experiment whose power is more than sufficient to resolve the tension. Generally though, we expect that future experiments will restrict us to regions of parameter space where we can more accurately model the observables as linear in parameters, and thus where our blinding transformation's performance improves.

For conciseness, we will refer to this as the 3 × 2pt analysis, so named because in it three types of 2PCF are used as summary statistics. These 2PCFs are galaxy position–position, shear–shear, and position–shear angular two-point correlations measured from DES galaxy catalogues. Additionally, we will use Year 1 (Y1) to refer to the analysis of the first year of DES data, which covers a footprint of roughly 1300 deg^2 and for which results are reported in Abbott et al. (2018), and Year 3 (Y3) to refer to the ongoing analysis of the first three years of data, which will cover the full 5000 deg^2 footprint at a similar depth.

Below, we first briefly motivate the need for blinding in DES combined-probe analyses (Section 4.1) before describing the 3 × 2pt data and analysis pipeline in Sections 4.2 and 4.3. Then, Section 4.4 introduces our procedure for testing the performance of the 2PCF blinding transformation. Sections 4.5 and 4.6 present the results.

4.1 The need for blinding in DES analyses

Two of the most powerful ways DES data can test the ΛCDM model are via the constraints it can place on the dark energy equation-of-state parameter w and on the amplitude of matter density fluctuations σ_8. The equation-of-state parameter describes the ratio of pressure and density of a fluid description of dark energy in wCDM, an extension of ΛCDM, which describes dark energy as a fluid. If dark energy behaves as a cosmological constant (as in ΛCDM), this parameter will take the value $w = -1$, while $w \neq -1$ means that the dark energy density evolves with time. The matter density fluctuation amplitude σ_8 is of interest because it and $\Omega_{\rm m}$ are the mostly precisely constrained parameters for DES' measurements of structure in the $z \lesssim 1$ Universe. Comparing DES constraints in the σ_8–$\Omega_{\rm m}$ plane to those extrapolated under ΛCDM from *Planck* CMB measurements thus tests the ability of ΛCDM to describe the evolution of the large-scale properties of the universe from early and late times. Given these tests, whether or not DES observables are consistent with the special value $w = -1$ in wCDM parameter space or with *Planck* $\Omega_{\rm m}$–σ_8 results in ΛCDM are questions of particular interest for DES analyses.

The Y1-3 × 2pt wCDM constraints are consistent with $w = -1$, while the ΛCDM results showed a suggestive offset from *Planck* in the $\Omega_{\rm m}$–σ_8 plane, with the DES preferring a lower value of σ_8.[4] As the Y3-3 × 2pt analysis will use three times the sky area, that increased statistical power will cause DES constraints to tighten, and the community will be closely watching how the results compare to $w = -1$ and to the *Planck* $\Omega_{\rm m}$–σ_8 constraints. Thus, the parameters that we are particularly interested in blinding are w and σ_8. In the tests below, for simplicity, we focus on blinding transformations shifting only those two parameters. This is sufficient to blind the DES analysis, but we do note that one could easily and reasonably adjust the transformation to include shifts in any other parameter(s) – $\Omega_{\rm m}$, for example – in order to further confound the DES-*Planck* comparison.

4.2 Observable vector and likelihood

Because the DES Y3-3 × 2pt pipeline was not finalized when the blinding investigations presented in this paper were conducted, we

[4]In Abbott et al. (2018), this offset is reported to be statistically insignificant, though exactly how such a tension should be quantified is a subject of some discussion (Handley & Lemos 2019).

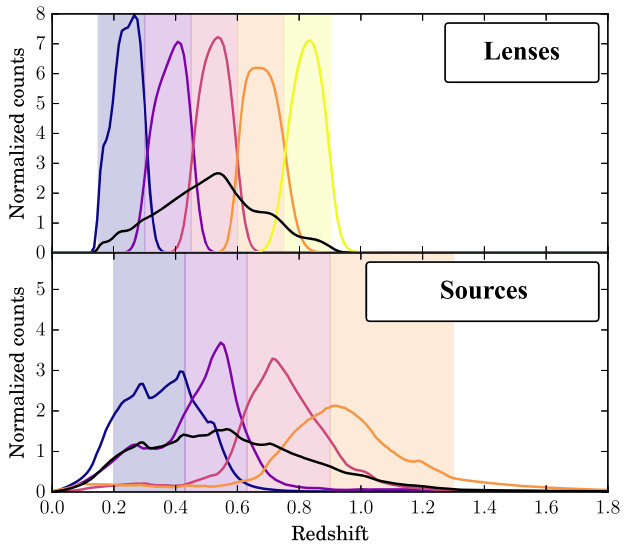

Figure 2. The $n(z)$ redshift distributions for lens and source galaxies in the DES Y1-3 × 2pt analysis from Abbott et al. (2018). The vertical coloured bands show the nominal redshift range of each bin, while the lines show the estimated true redshift distribution when galaxies are binned in photometric redshift. The black lines show the unbinned total distribution. We adopt these same redshift distributions for our Y3 blinding tests.

approximate the Y3 observable vector and likelihood by using the Y1 modelling choices, prior ranges, and scale cuts. This section will briefly describe them, but we refer the reader to Abbott et al. (2018) and Krause et al. (2017) for a more detailed description of the associated measurements and calculations.

The DES 3 × 2pt analysis is based on observations of two populations of galaxies. Positions are measured for a set of lens galaxies which have been selected to have small photometric redshift (photo-z) errors and which have been carefully checked for residual systematics. In the Y1 analysis, this population consisted of 650 000 bright red-sequence galaxies which are selected as part of the redMaGiC catalogue (Elvin-Poole et al. 2018). Their redshift distribution is shown in the upper panel of Fig. 2. Cosmic shears are measured from a larger population of source galaxies. For the Y1 analysis the source galaxies included 26 million objects selected from the Y1 Gold catalogue (Drlica-Wagner et al. 2018) and their shapes are measured as described in Zuntz et al. (2018). The lens (source) galaxies are divided into five (four) redshift bins, respectively; see Fig. 2.

The 3 × 2pt observable vector consists of three kinds of 2PCF measured from the lens and source catalogues. The galaxy–galaxy correlations are measured as autocorrelations within each lens bin, producing a set of functions $w^i(\theta)$ for $i = 1$–5. Shear-shear correlations are measured for all auto and cross-correlations of the source bins, producing functions $\xi_+^{ij}(\theta)$ and $\xi_-^{ij}(\theta)$ for $i = 1$–4 and $1 \leq j \leq i$. The galaxy-shear cross-correlations are measured between all combinations of the five lens bins and four source bins, producing $\gamma_t^{ij}(\theta)$ for $i = 1$–5 and $j = 1$–4. All of these 2PCFs are measured for twenty logarithmically spaced angular bins between 2.5 and 250 arcmin. Further scale cuts are applied in order to prevent modeling uncertainties associated with non-linear structure formation, baryonic physics, and other small-scale effects from biasing the final cosmological results (Krause et al. 2017). The resulting 3 × 2pt observable vector has 457 entries.

The likelihood of the DES 3 × 2pt observable vector is modelled as a multivariate Gaussian. Its covariance C_d has significant off-diagonal contributions, since many elements of the observable vector can share dependence on the realization of the mass and galaxy distributions in the survey volume (also known as sample variance). We adapt the covariance matrix that was previously analytically computed for the Y1 analysis as described in Krause et al. (2017) using COSMOLIKE (Krause & Eifler 2017). To approximate the Y3 covariance, we simply scale the Y1 covariance by a factor of 0.27 = 1350/5000 to account for Y3's increased survey area. This survey-area scaling correctly modifies the Gaussian parts of the covariance, but it does not properly scale the non-Gaussian contributions (Joachimi, Schneider & Eifler 2007). Thus, this is only a rough approximation for the Y3 covariance, but it is sufficient for our testing purposes. Though, in principle, the 3 × 2pt covariance depends on the model parameters, it has been shown (Eifler, Schneider & Hartlap 2009) that the covariance's cosmology dependence can be neglected without significantly affecting parameter constraints. In the DES Y1-3 × 2pt analysis and in this work, we do not vary the data covariance matrix when performing parameter searches.

4.3 Modelling

The parameter estimation procedure for the DES Y1-3 × 2pt analysis, and therefore also our simulated Y3-3 × 2pt analysis, involves a search over 27 free parameters for wCDM. This includes seven cosmological parameters[5] (Ω_m, σ_8, w, h, Ω_b, n_s, and $\Omega_\nu h^2$) and 20 nuisance parameters used to account for various systematic uncertainties. These nuisance parameters include a constant linear galaxy bias b_i for each of the five lens redshift bins. Additional nuisance parameters are introduced to model the effects of uncertainties in photo-z estimation: for each lens and source bin, we introduce a parameter Δz_i that describes a redshift offset of that bin's $n(z)$ distribution. To model shear calibration, we assign one multiplicative shear calibration parameter m_i per source galaxy bin. Following the Y1 analysis, we impose tight Gaussian external priors on all shear calibration and photo-z nuisance parameters. The last set of nuisance parameters model how intrinsic (as opposed to lensing-induced) alignments between galaxy shapes affect their observed 2PCF. We use a linear alignment model with parameters A_{IA}, α_{IA}, and $z_0^{(IA)}$. The calculations we use to compute predictions for the 3 × 2pt observable vector, given this set of model parameters, are described in Appendix B and in more detail in Krause et al. (2017).

The fiducial values for all of these nuisance parameters, as well as cosmological parameters, are shown in Table 1, as are the Gaussian priors used for the photo-z shifts Δz_i and shear calibrations m_i. During parameter estimation, the number of neutrinos $N_{\mathrm{massive}\nu}$ and $N_{\mathrm{massless}\nu}$ (chosen to sum to the standard model effective number of neutrinos N_{eff}), the optical depth of the CMB τ, and $z_0^{(IA)}$ are fixed, while the rest of the parameters shown in the table were varied with flat priors.

4.4 Evaluating performance for DES blinding

Our goal is to test the performance of 2PCF-based blinding for Y3-3 × 2pt. We do this by analysing an ensemble of 100 noiseless

[5]Note that while for the purposes of this blinding study we sample over σ_8 as an input model parameter, Abbott et al. (2018) and other DES analyses typically sample over A_s and measure σ_8 as a derived parameter.

Table 1. Fiducial values and prior ranges for the DES 3 × 2pt analysis pipeline studied in this work along with the range from which parameters are drawn for the blinding tests performed in this study.

			Parameter estimation		Distribution for blinding tests		
	Parameter	$\Theta_{\rm ref}$	Search bounds	Prior	$\Theta_{\rm ref}+\Delta\Theta$	$\Theta_{\rm obs}$ [fid.]	$\Theta_{\rm obs}$ [Nuis.]C
Cosmology parameters	σ_8	0.834	[0.1, 2.0]	Flat	[0.734, 0.954]	Y3 Fisher	Matches fid.
	w	−1.0	[−2.0, −0.0]	Flat	[−1.5, 0.5]	Y3 Fisher	Matches fid.
	$\Omega_{\rm m}$	0.295	[0.1, 2.0]	Flat	–	Y3 Fisher	Matches fid.
	h	0.6882	[0.2,1.0]	Flat	–	Y3 Fisher	Matches fid.
	$\Omega_{\rm b}$	0.0468	[0.03, 0.07]	Flat	–	–	–
	$n_{\rm s}$	0.9676	[0.87, 1.07]	Flat	–	–	–
	$\Omega_\nu h^2$	6.155×10^{-4}	[0.0006, 0.01]	Flat	–	–	[0.0006,0.00322]
	$N_{\rm massive}\nu$	3	–	Fixed	–	–	–
	$N_{\rm massless}\nu$	0.046	–	Fixed	–	–	–
	τ	0.08	–	Fixed	–	–	–
Lens galaxy bias	b_1	1.45	[0.8, 2.5]	Flat	–	Y3 Fisher	Full prior
	b_2	1.55	[0.8, 2.5]	Flat	–	Y3 Fisher	Full prior
	b_3	1.65	[0.8, 2.5]	Flat	–	Y3 Fisher	Full prior
	b_4	1.8	[0.8, 2.5]	Flat	–	Y3 Fisher	Full prior
	b_5	2.0	[0.8, 2.5]	Flat	–	Y3 Fisher	Full prior
Shear calib.	m_{1-4}	0.012	[−0.1, 0.1]	$\mathcal{N}(0.012, 0.023)$	–	–	[−0.57,0.081]
Intrinsic alignments	$A_{\rm IA}$	0.0	[−5, 5]	Flat	–	–	[0,1]
	$\alpha_{\rm IA}$	0.0	[−5, 5]	Flat	–	–	[−4, 4]
	$z_0^{\rm (IA)}$	0.62	–	Fixed	–	–	–
Source galaxy photo-z bias	$\Delta z_1^{\rm source}$	−0.002	[−0.1,0.1]	$\mathcal{N}(-0.001, 0.016)$	–	–	[−0.05, 0.046]
	$\Delta z_2^{\rm source}$	−0.015	[−0.1,0.1]	$\mathcal{N}(-0.019, 0.013)$	–	–	[−0.054, 0.024]
	$\Delta z_3^{\rm source}$	0.007	[−0.1,0.1]	$\mathcal{N}(0.009, 0.013)$	–	–	[−0.026, 0.04]
	$\Delta z_4^{\rm source}$	0.018	[−0.1,0.1]	$\mathcal{N}(-0.018, 0.022)$	–	–	[−0.048, 0.84]
Lens galaxy photo-z bias	$\Delta z_1^{\rm lens}$	0.002	[−0.05, 0.05]	$\mathcal{N}(0.008, 0.007)$	–	–	[−0.022, 0.026]
	$\Delta z_2^{\rm lens}$	0.001	[−0.05, 0.05]	$\mathcal{N}(-0.005, 0.007)$	–	–	[−0.020, 0.022]
	$\Delta z_3^{\rm lens}$	0.003	[−0.05, 0.05]	$\mathcal{N}(0.006, 0.006)$	–	–	[−0.018, 0.024]
	$\Delta z_4^{\rm lens}$	0.0	[−0.05, 0.05]	$\mathcal{N}(0.0, 0.01)$	–	–	[−0.03, 0.03]
	$\Delta z_5^{\rm lens}$	0.0	[−0.05, 0.05]	$\mathcal{N}(0.0, 0.01)$	–	–	[−0.03, 0.03]

Notes. Fiducial values and priors are chosen to be similar to the settings for the DES Y1-3 × 2pt analysis (with the exception of w, which uses a wider range here than in Abbott et al. 2018), and the fact that for these tests, we search over σ_8 rather than $A_{\rm s}$. The columns labelled 'Distribution for blinding tests' display the range from which parameters are drawn to create the realizations of synthetic data and blinding factors used for the blinding tests and are described in Sections 4.4.1 and 4.4.1. The 'Y3 Fisher' label in the $\Theta_{\rm obs}$ column refers a multivariate Gaussian, distribution with the parameter covariance estimated using a Fisher forecast, centred on the fiducial parameter values. For $\Omega_\nu h^2$, the upper bounds of its prior range and the upper bound of its 'Nuis.' range correspond to $\sum m_\nu = 0.93$ and 0.3 eV, respectively.

synthetic 3 × 2pt observable vectors. For each realization, we draw $\Delta\Theta$ (which determines the blinding transformation) and $\Theta_{\rm obs}$ (which determines the 'truth') from probability distributions centred on a reference cosmology $\Theta_{\rm ref}$. That reference cosmology is fixed to the fiducial parameter values listed in Table 1. The synthetic 'observed' data $\hat{\boldsymbol{d}}$ are then generated by computing a theory prediction for the 3 × 2pt observable vector at parameters $\Theta_{\rm obs}$, so

$$\hat{d}_i = d_i(\Theta_{\rm obs}). \tag{17}$$

That data are then transformed according to equation (7) to produce a blinded observable vector

$$B(\hat{d}_i) = \hat{d}_i + d_i\left(\Theta_{\rm ref}+\Delta\Theta\right) - d_i\left(\Theta_{\rm ref}\right). \tag{18}$$

We then search for parameters $\Theta_{\rm bl}$ and $\Theta_{\rm unbl}$ that maximize the likelihood (minimize χ^2) for the blinded and unblinded data. The change $\Delta\chi^2$ induced by applying the blinding transformation to the data will be our measure of success for the blinding transformation, since the blinded data should look compatible with some model in the parameter space. For select realizations, we additionally study the impact of blinding on the parameter estimation posteriors, as is shown in Fig. 4. The following subsections describe the technical details of this procedure.

4.4.1 Parameter selection: fiducial test

The $\Delta\Theta$ distribution is chosen to satisfy criterion I from Section 2.1 above, while $\Theta_{\rm obs}$ will be drawn from a range that reasonably reflects potential offsets between the true Y3-3 × 2pt cosmology and $\Theta_{\rm ref}$. We blind the two cosmological parameters that are at the greatest risk of experimenters' bias in the DES 3 × 2pt analysis: the amplitude of matter clustering σ_8 and the dark energy equation-of-state parameter w. We draw $\Delta\sigma_8$ from a flat distribution centred on zero with bounds $-0.12 < \Delta\sigma_8 < +0.12$ chosen to be roughly equal to the $\pm3\sigma$ errors expected from Y3-3 × 2pt.[6] We draw w from a flat distribution $-0.5 < \Delta w < +0.5$, chosen to span half of the flat prior being used for parameter estimation. All other parameters have no input blinding shift.

[6]For comparison, the ΛCDM constraints on σ_8 reported in Abbott et al. (2018) are $0.817^{+0.045}_{-0.056}$.

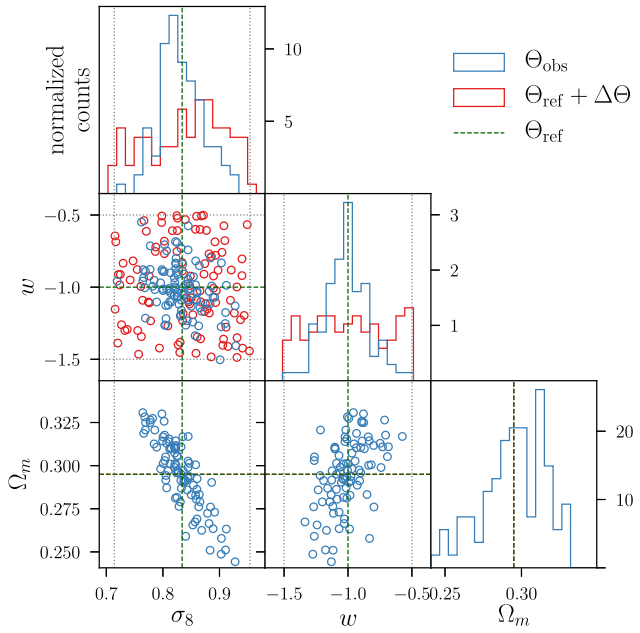

Figure 3. Values (off-diagonal panels) and the distribution (diagonal panels) of parameters σ_8, w, and $\Omega_{\rm m}$ in the 100 realizations used for the fiducial test of our blinding procedure. The blue circles are the observed (that is true, unblinded) parameter values. The red circles are the shifted values used to blind the data. The dashed lines denote the reference model used for blinding.

For our fiducial test, we vary $\Theta_{\rm obs}$ over a subset of the 27 3 × 2pt wCDM parameters: σ_8, w, $\Omega_{\rm m}$, h, and the five lens galaxy bias parameters b_i for $i \in \{1, \ldots, 5\}$. It is drawn from a truncated multivariate Gaussian distribution in those parameters centred on $\Theta_{\rm ref}$. The covariance of that distribution is obtained from Fisher forecast for our Y3-3 × 2pt pipeline fixing all unvaried parameters. Any realization for which $\Theta_{\rm obs}$ falls outside of the flat prior ranges in Table 1 is discarded and redrawn. Fig. 3 shows the resulting 100 realizations of $\Theta_{\rm ref} + \Delta\Theta$ and $\Theta_{\rm obs}$ for a subset of parameters.

4.4.2 $\Delta\chi^2$ threshold

The blinded data should appear fully consistent with having been generated by some model within $\mathcal{M}$ (here, wCDM within the priors shown in Table 1), if and only if the unblinded data are consistent with the true model. Since χ^2 is our measure of data-model consistency, this means that we would ideally like the blinding of the data to result in $\Delta\chi^2 = 0$. A finite $\Delta\chi^2$ is tolerable, however, if it is smaller than the expected statistical variation in the unblinded χ^2, i.e. $\Delta\chi^2$ will not influence acceptance or rejection of the data.

We choose $\Delta\chi^2 = 30$ as a threshold for acceptable contributions from blinding to the error in the fit. This is within a few per cent of the standard deviation $\sigma_{\chi^2} \approx \sqrt{2\nu}$ for the number of degrees of freedom ν being considered in our analyses. For comparison, in the DES Y1-3 × 2pt analysis of Abbott et al. (2018), an unblinding criterion was that $\chi^2/\nu < 1.4$. Our $\Delta\chi^2 = 30$ threshold corresponds to $\Delta\chi^2/\nu = 0.07$.

4.4.3 Finding the maximum likelihood

To find the best-fitting parameters, we use the MAXLIKE sampler in COSMOSIS, which is a wrapper for the `scipy.optimize.minimize` function using the Nelder–Mead Simplex algorithm (Nelder & Mead 1965). This routine can fail to find the true maximum likelihood in high-dimensional spaces, biasing χ^2 high. For our noise-free tests, we know $\chi^2 = 0$ for the unblinded data, hence the measured $\Delta\chi^2$ values are strict upper bounds on the true contributions of blinding to χ^2. To more accurately characterize this bound, we re-run the χ^2 minimization search for all realizations with $\Delta\chi^2 > 30$ using the MULTINEST sampler[7] (Feroz & Hobson 2008; Feroz, Hobson & Bridges 2009; Feroz et al. 2013) (as implemented in COSMOSIS). MULTINEST is more computationally costly than MAXLIKE, but because it more thoroughly explores the parameter space, it is less susceptible to getting stuck in local minima. We perform the MULTINEST searches over the full 27-dimension wCDM parameter space using the same low-resolution settings used for DES Y1-3 × 2pt exploratory studies[8] and substitute the MAXLIKE results with those from MULTINEST results in cases where the minimum χ^2 reported from MULTINEST is smaller.[9]

4.5 Results for fiducial test

Fig. 4 shows an example of how the parameter contours shift in response to a large blinding shift of ($\Delta\sigma_8 = -0.117$, $\Delta w = +0.314$). For this realization, the change in goodness of fit is small, $\Delta\chi^2 = 1.86$, and we can see that the location of the 68 and 95 per cent confidence contours change in a way consistent with the input blinding shift.

More quantitatively, Fig. 5 contains the primary results for our test of the 2PCF-based blinding for the DES Y3-3 × 2pt analysis. It shows a histogram of the $\Delta\chi^2$ values for the 100 blinded and unblinded pairs of synthetic observable vectors analysed. We see that in all realizations $\Delta\chi^2$ is well below our $\Delta\chi^2 = 30$ cut-off. This means that the blinded data are indistinguishable from unblinded data at different input parameters.

In Section 3.3, we suggested that the performance of the blinding transformation should worsen (higher $\Delta\chi^2$) when both the blinding shift $\Delta\Theta$ and the truth shift $\Theta_{\rm obs} - \Theta_{\rm ref}$ become significantly non-zero. Fig. 6 plots the $\Delta\chi^2$ of each fiducial realization (as the colour) versus the size of the blinding shift and the truth shift. To quantify the size of these shifts, we extract a parameter covariance C from a fiducial MULTINEST chain and use it to compute distances in parameter space:

$$|\Theta_{\rm obs} - \Theta_{\rm ref}|/\sigma_\Theta = \sqrt{(\Theta_{\rm obs} - \Theta_{\rm ref})^T C^{-1} (\Theta_{\rm obs} - \Theta_{\rm ref})}, \quad (19)$$

$$|\Delta\Theta|/\sigma_\Theta = \sqrt{(\Delta\Theta)^T C^{-1} (\Delta\Theta)}. \quad (20)$$

For the blinding shift, we compute this distance in the marginalized two-dimensional $\sigma_8 - w$ plane (that is, we use a 2 × 2 C containing only the entries for those parameters). We evaluate the truth shift in the full 27-dimensional parameter space. For ease of interpretation, these distances are then converted into an equivalent deviation for a one-dimensional Gaussian normal distribution by equating the

[7] `ccpforge.cse.rl.ac.uk/gf/project/multinest/`.
[8] These MULTINEST settings are: 250 live points, efficiency 0.8, tolerance 0.1.
[9] Because MULTINEST is designed to map the posterior distribution rather than find the best-fitting point in parameter space, its accuracy will be limited by the density of samples in the high likelihood region. Based on MULTINEST fits to unblinded observable vectors (where we know the true minimum is $\chi^2 = 0$), we estimate that the MULTINEST results tend to overestimate the minimum χ^2 by ∼5.

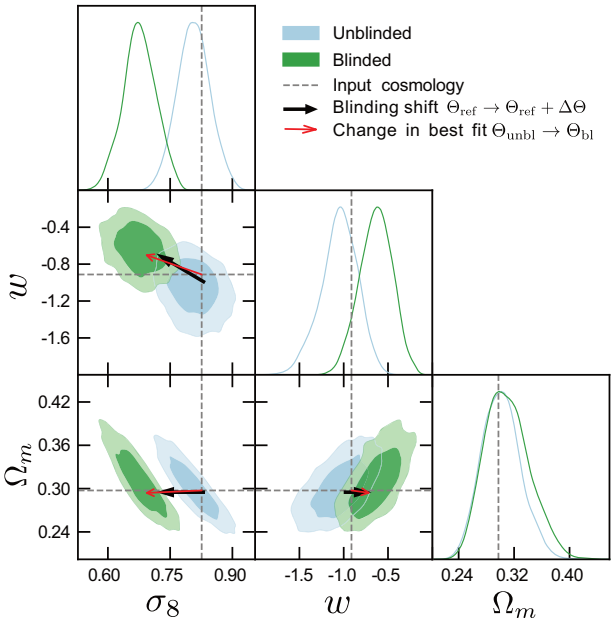

Figure 4. The 68 and 95 per cent confidence intervals for blinded (green) and unblinded (blue) synthetic DES Y3-3 × 2pt data. As an illustrative example, results are shown for a realization of the fiducial blinding test with a large input blinding shift of ($\Delta\sigma_8 = -0.117$, $\Delta w = +0.314$) and a low $\Delta\chi^2 = 1.86$. Dashed grey lines show the input parameters used to simulate the unblinded data $\Theta_{\rm obs} = (\sigma_8 = 0.826,\ w = -0.912,\ \Omega_{\rm m} - 0.29)$. The black, thick arrow shows the points from $\Theta_{\rm ref}$ to $\Theta_{\rm ref} + \Delta\Theta$ to show the input blinding shift, and the red, thin arrow points from $\Theta_{\rm unbl} = \Theta_{\rm obs}$ to $\Theta_{\rm bl}$ to show the change in best-fitting parameters.

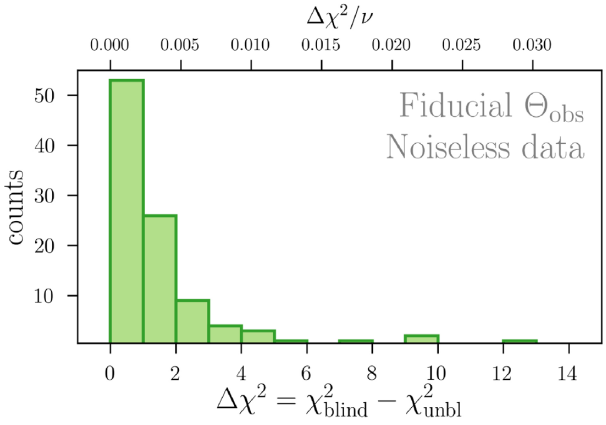

Figure 5. Fiducial blinding test results for $\Delta\chi^2$. The top axis shows $\Delta\chi^2$ in units of the degrees of freedom associated with the DES Y3-3 × 2pt wCDM analysis, $\nu = 430$.

probability-to-exceed value: we solve for N such that

$$P(\chi^2 > N^2, \nu = 1) = P(\chi^2 > (|\Delta\Theta|/\sigma_\Theta), \nu = 2) \quad (21)$$

and likewise for the truth shift (but with $\nu = 27$ on the right). For example, a truth shift of $|\Theta_{\rm obs} - \Theta_{\rm ref}|/\sigma_\Theta = \sqrt{30} \simeq 5.5$ that corresponds to $\chi^2 = 30$ in 27-dimensional parameter space has the same probability-to-exceed as a $\chi^2 = 1$ signal in a one-dimensional Gaussian, so we would plot this as a '1σ' blinding shift in the metric of the parameter covariance matrix.

Fig. 6 confirms the behaviour derived in Section 3.3 that the larger $\Delta\chi^2$ values appear only when both $\Theta_{\rm ref} + \Delta\Theta$ and $\Theta_{\rm obs}$ move

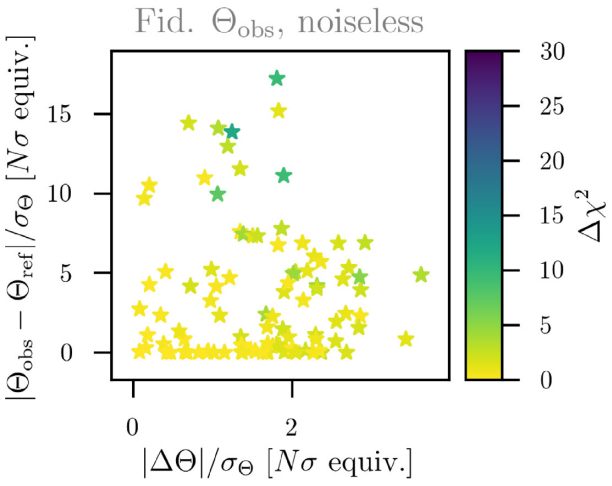

Figure 6. Dependence of $\Delta\chi^2$ on the magnitude of the truth shift $|\Theta_{\rm obs} - \Theta_{\rm ref}|/\sigma_\Theta$ (vertical axis) and of the blinding shift $|\Delta\Theta|/\sigma_\Theta$ (horizontal axis) for the observable vectors generated using the fiducial 100 realizations of true cosmology $\Theta_{\rm obs}$. These distances are evaluated in the two-dimensional $\sigma_8 - w$ plane for the blinding shift and in the full 27-dimensional parameter space for the truth shift. In both cases the distance is scaled according to expected parameter uncertainties and are shown in units of the number of standard deviations for a one-dimensional Gaussian distribution with equivalent probability-to-exceed. The colours of the points represent the $\Delta\chi^2$ value of those realizations.

significantly away from $\Theta_{\rm ref}$ under the metric of the experimental posterior. This trend is non-monotonic because the performance of blinding depends somewhat on the direction in parameter space of the vector $\Theta_{\rm obs} - \Theta_{\rm ref}$ in addition to its magnitude. Furthermore, there is noise in our $\Delta\chi^2$ evaluation from imperfect optimisation.

Fig. 7 shows the relationship between the input blinding shifts in σ_8 and w and the resulting shift in their best-fitting values. We generally find the behaviour expected from the leading-order analysis, which is for $\Theta_{\rm bl} - \Theta_{\rm unbl}$ to be roughly equal to the parameter shift $\Delta\Theta$ used to generate the blinding factors, although with some deviation that grows roughly linearly with $\Delta\Theta$. The scatter is expected because unplotted parameters, as well as the truth shift $\Theta_{\rm obs} - \Theta_{\rm ref}$, also influence the deviation. The fact that the range of the points along the veritical axes of Fig. 7 is comparable to that along the horizontal axis does confirm that this blinding transformation is capable of altering the experiment's results enough to overcome experimenters' potential biases. Having satisfied all of the desired criteria described in Section 2, we consider the blinding transformation of equation (7) to be successful for this fiducial analysis.

Fig. 7 can give us additional insight into the performance of our blinding transformation. In that figure, the colour of each point corresponds to the $\Delta\chi^2$ value for that realization. We note that the strong outlier points are also the points with largest $\Delta\chi^2$. This is the expected behaviour. We saw in Fig. 6 that these high-$\Delta\chi^2$ realizations have large truth shifts, and we expect that larger truth shifts should induce larger differences in $\Theta_{\rm bl} - \Theta_{\rm obs}$, mediated by the non-linear portions of the data model as indicated by equation (A6). Those differences are expected to occur for all parameters, not just those we choose to blind. We confirm from our MAXLIKE results that, for such outlier realizations of $\Theta_{\rm obs}$, blinding induces large changes in the best-fitting values of several nominally unblinded parameters.

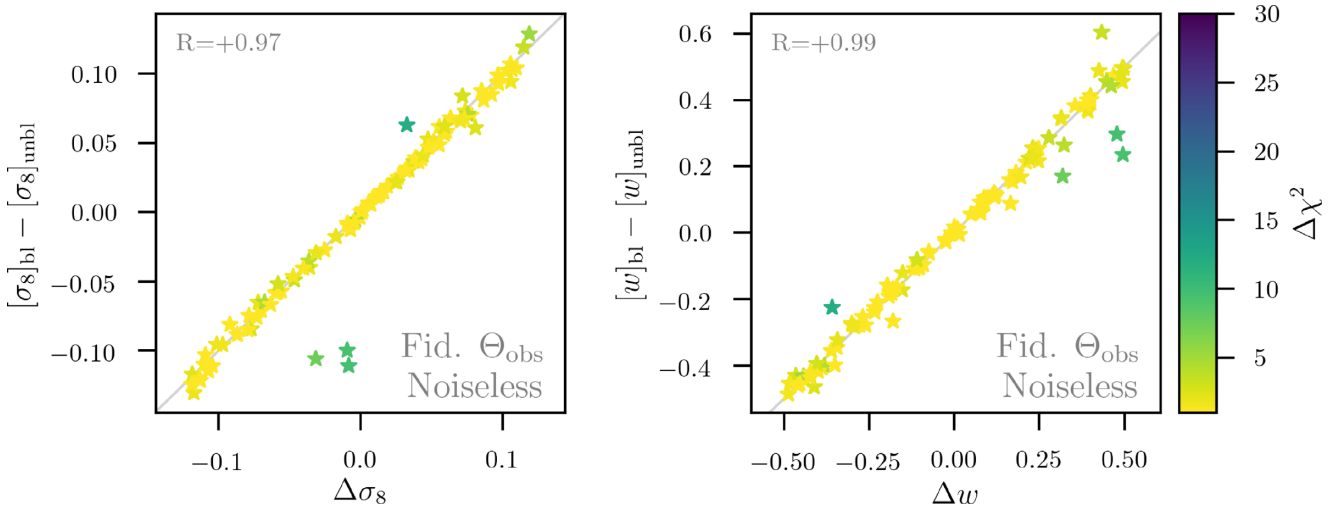

Figure 7. Relationship between the input blinding shifts (horizontal axes) in σ_8 (left-hand panel) and w (right-hand panel) and the output shifts in best-fitting values of those same parameters. The colour scale shows $\Delta\chi^2$ and the diagonal grey line shows the case where the output shifts are exactly equal to the input shifts. Both plots show data from the same 100 realizations for noiseless simulated data.

To further investigate this, we additionally ran MULTINEST chains for those realizations. We found that in two of these realizations the 68 or 95 per cent confidence intervals of posterior are pushed into the hard-prior boundary for some of the galaxy bias parameters. Such behaviour could be problematic, since checking that posterior distributions have not hit prior bounds is a standard part of data validation, and the blinding could trigger a false alarm for this. This issue not necessarily prohibitive: Since we know that this kind of unpredictable shift is occurring for realizations with large truth offsets $\Theta_{\rm obs} - \Theta_{\rm ref}$, one could imagine workarounds in the blinding procedure to account for, or protect against, this possibility. For example, the collaboration could ask a single member to confirm whether the collision with the prior bound is, in fact, due to the blinding shift or, as discussed in Section 5 below, one could make use of several distinct blinding shifts defined at different $\Theta_{\rm ref}$ values.

4.5.1 *Impact of noise*

As an additional check, we perform this analysis on a version of these same 100 synthetic observable vectors with Gaussian noise added using a Cholesky decomposition of the covariance C_d. Results from this test are shown in Appendix C. Compared to the noiseless results shown above, there is slightly more scatter in the input-versus-output parameter shift relationship, and in the relationship between $\Delta\chi^2$ and the various input parameters. These differences are expected and do not change the conclusions that the blinding transformation is effective.

4.6 Follow-up test: varying nuisance parameters

As a stronger test of 2PCF-based blinding for the DES Y3-3 × 2pt analysis, we analyse a second set of synthetic observable vectors for which more parameters of $\Theta_{\rm obs}$ are allowed to deviate from $\Theta_{\rm ref}$. This is a more rigorous test because it allows for larger differences between the truth model and the reference model; recall that we noted in Section 3.3 that we expect the performance of the blinding method to degrade as the magnitude of the difference $\Theta_{\rm obs} - \Theta_{\rm ref}$ increases. We use the same 100 blinding factors as in the fiducial test. For the 100 realizations of $\Theta_{\rm obs}$, the values of σ_8, w, $\Omega_{\rm m}$, h match those used in the fiducial test, but we additionally vary $\Omega_\nu h^2$, reselect values of galaxy bias b_{1-5}, and vary all remaining nuisance parameters over flat probability distributions with ranges shown in the rightmost column of Table 1. For the nuisance parameters with Gaussian priors in the 3 × 2pt analysis, the ranges are $\pm 3\sigma$ for that prior. The galaxy bias parameters are drawn from their full prior range, while the ranges for neutrino mass and intrinsic alignment parameters are chosen by eye based on the DES Y1-3 × 2pt posteriors in Abbott et al. (2018). In the text and plots below, we will refer to results from this test as the 'nuisance' test (as opposed to fiducial test).

After searching for best-fitting parameters[10] on the pairs of blinded and unblinded observable vectors of the nuisance test, we can study the same results discussed in Section 4.5. Fig. 8 shows a histogram of the $\Delta\chi^2$ values. Although the majority of realizations remain below $\Delta\chi^2 < 30$, the distribution is broader than in the fiducial test and four realizations exceed that threshold.

To understand the nature of the failures, we examine how $\Delta\chi^2$ depends on the blinding shifts and truth shifts of each realization. Fig. 9 plots the distribution of $\Delta\chi^2$ values for the nuisance test in the space of blinding shifts applied to σ_8 and w. Here it is apparent that larger blinding shifts, particularly those that reduce σ_8, are associated with higher $\Delta\chi^2$. A more revealing picture is given in Fig. 10, where it is clear that the unacceptable $\Delta\chi^2$ trials are those with larger blinding shifts *and* extremely large distances between $\Theta_{\rm obs}$ and $\Theta_{\rm ref}$ (under the posterior parameter metric). Note that change of scale on the y-axis from Fig. 6.

[10]When we run MULTINEST for select realizations of the nuisance test, we fit over the likelihood rather than the posterior. We do this because for many realizations, the prior is very small at the true cosmology $\Theta_{\rm obs}$, causing significant differences between the maxima of the posterior and likelihood. This procedure is specific to the synthetic data study presented here where we are purposefully drawing $\Theta_{\rm obs}$ from a wide range of nuisance parameter values, and where we are trying to minimize χ^2 rather than maximize the posterior. For the analysis of real data, nuisance parameter priors should reflect our knowledge of what values they are likely to take, and one should maximize the posterior.

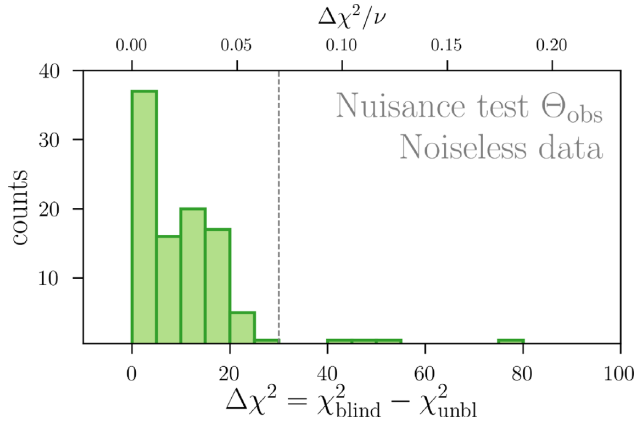

Figure 8. Same as Fig. 5, but for the 'nuisance test' ensemble of unblinded observable vectors: histogram of $\Delta\chi^2$ values. The vertical dashed line shows our $\Delta\chi^2 = 30$ threshold, which is now exceeded by 4 of the 100 trials.

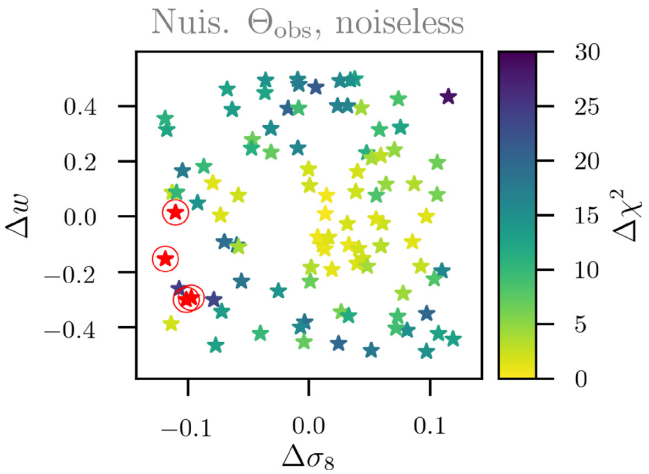

Figure 9. For the 'nuisance test' ensemble, we plot the $\Delta\chi^2$ induced by the blinding process as the colour of each point, versus the blinding shifts applied to the two cosmological parameters of greatest interest. Points with $\Delta\chi^2 > 30$ are shown in red and are circled.

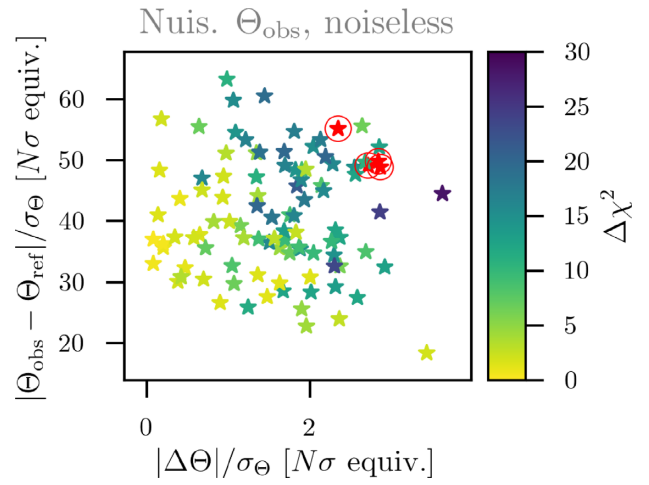

Figure 10. Same as Fig. 6, but for the 'nuisance test' ensemble of unblinded observable vectors: dependence of $\Delta\chi^2$ on the parameter-space distances associated with $\Theta_{\rm obs} - \Theta_{\rm ref}$ and $\Delta\Theta$. Circled points shown in red are realizations with $\Delta\chi^2 > 30$.

In interpreting these results, we emphasize that this is by design a conservative test of the blinding transformation. The fact that the true cosmologies $\Theta_{\rm obs}$ are drawn from a flat probability distribution in a large number of nuisance parameters means that the distances shown on the vertical axis of Fig. 10 are very large relative to projected Y3-3 × 2pt uncertainties. In other words, all of the realizations studied for the nuisance test have truth parameters that are collectively highly unlikely under the priors that have been assigned to them,[11] and even more unlikely under the expected posterior probability. To put this in perspective, note that a more realistic simulation of the blinding transformation's performance could be done by drawing $\Theta_{\rm obs}$ from the DES Y1-3 × 2pt posterior probability distribution, which might be considered an appropriate prior for Y3. The resulting ensemble of $\Theta_{\rm obs} - \Theta_{\rm ref}$ separations would be significantly smaller than those used here, we would expect better performance in our $\Delta\chi^2$ test. However, especially given the limited number (100) of realizations, sampling $\Theta_{\rm obs}$ from the Y1 posterior would result in few realizations with large $\Theta_{\rm obs} - \Theta_{\rm ref}$, which might test the limits of this approach to blinding. In contrast, our nuisance test's flat probability distributions and resulting extremely large truth shifts let us probe regions of parameter space where the blinding transformation breaks down. We then check that those regions are highly unlikely for the application we are considering. We conclude that, although the nuisance test has failed blinding Criterion 2 of Section 2.2 in 4 per cent of our trials, this only occurs for realizations with true parameters that are much farther from our reference parameters than they would realistically be for DES Y3-3 × 2pt.

Fig. 11 shows how blinding shifts relate to output shifts in the best-fitting parameters for the nuisance test. Compared to similar results from the fiducial test (Fig. 7), there is more scatter in this relationship and the high-$\Delta\chi^2$ points depart further from the trends. These behaviours are consistent with what we expect from expanding the $\Theta_{\rm obs}$ space. There is some curvature present in these relations at low w, which suggests that the model is extending to regions where the quadratic approximation assumed in Appendix A is inadequate. Nonetheless, we still confirm that the output parameter shifts are large enough to defeat observer prejudice.

5 DISCUSSION AND CONCLUSIONS

This paper presents and demonstrates an effective blinding strategy for multiprobe cosmological analyses at the summary-statistic level, with the principal requirement that such a strategy allows for robust validation checks (e.g. inspection of the blinded observable vector for obvious systematics) while hiding the cosmological-parameter values which are eventually to be inferred.

The blinding transformation is described in equation (7): One simply adds the difference between two theoretically-calculated observable vectors to the measured data. The 'blinding shift' is the prediction for a reference cosmology $\Theta_{\rm ref}$ subtracted from the prediction for shifted cosmology $\Theta_{\rm ref} + \Delta\Theta$. In the limit where the measurement noise is invariant and the summary statistics are linear functions of the parameters Θ, this transformation will generate blinded summary statistics that are completely indistinguishable from real data generated by the same experiment at an altered set

[11] The Gaussian nuisance prior probabilities π evaluated at input truth values $\Theta_{\rm obs}$ fall in the range $10^{-40} < \pi(\Theta_{\rm obs}) < 10^{-10}$ for all nuisance test realizations.

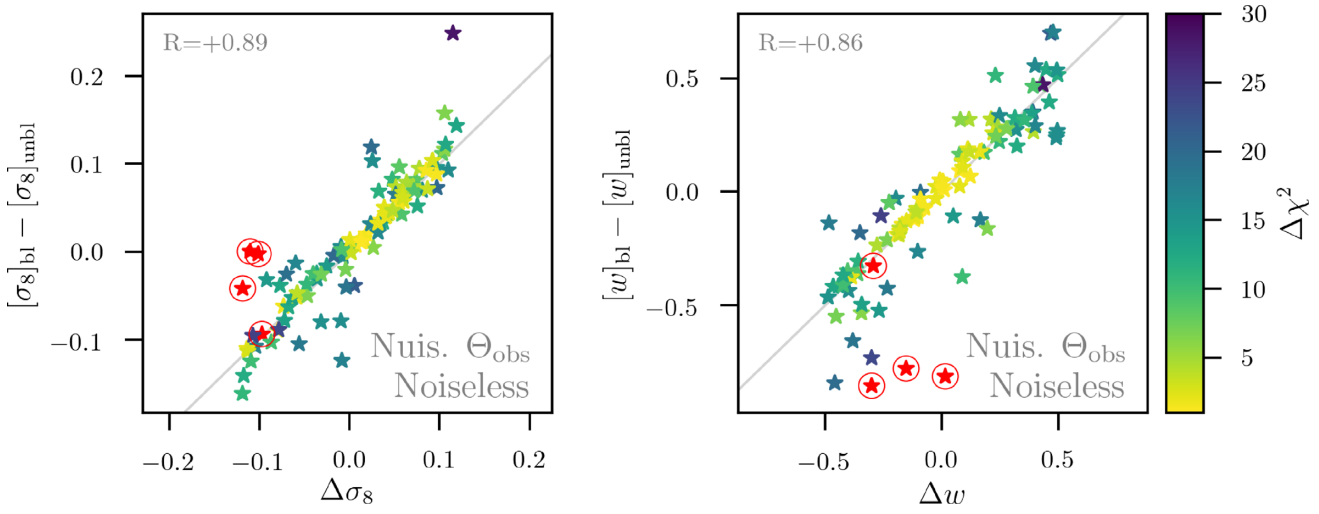

Figure 11. Same as Fig. 7 but for the 'nuisance test' ensemble of unblinded observable vectors: the relationship between input blinding shifts $\Delta\Theta$ and output shifts in best-fitting parameters $\Theta_{\rm bl} - \Theta_{\rm unbl}$. Realizations with $\Delta\chi^2 > 30$ are shown with circled red points. The grey lines correspond to equal input and output shifts.

of parameters $\Theta_{\rm obs} + \Delta\Theta$, i.e. a perfect blinding. This limit may or may not hold in the region spanned by the reference parameters, the shifted parameters, and the true-sky parameters $\Theta_{\rm obs}$. In practice, the model $d(\Theta)$ for the summary statistics will have non-linearity over the parameter range of interest, and one must check that the blinded data are close (in a χ^2 sense) to data that could be generated by *some* valid set of parameters $\Theta_{\rm bl}$. Therefore, most of this paper is devoted to verifying that this is the case for the forecasted DES Year 3 galaxy clustering and weak lensing, or Y3-3 × 2pt, analysis. We also check that this blinding transformation is capable of hiding the true results of the analysis, i.e. that this transformation can change the results enough so that that knowing $\Theta_{\rm bl}$ does not allow experimenters to know how the data compare to their prejudices.

These results serve both as a concrete example of the blinding transformation and as verification that it is reliable enough to use for the real DES Y3-3 × 2pt analysis. We focus on blinding shifts in σ_8 and w, and performed this test by finding the best-fitting parameters for blinded and unblinded versions of 100 realizations of noiseless synthetic measured observable vectors. (See Fig. 4 for an illustrative example.) In order to mimic the fact that the true cosmology will not match our fiducial $\Theta_{\rm ref}$, we vary the input parameters $\Theta_{\rm obs}$ for those simulations. For a proof-of-concept fiducial test, we select a subset of cosmologial parameters from a Fisher-forecast-based multivariate Gaussian distribution, and for a more conservative follow-up 'nuisance test', we additionally draw the values of nuisance parameters from flat probability distributions, resulting in very large $\Theta_{\rm obs} - \Theta_{\rm ref}$ offsets. The change in goodness-of-fit $\Delta\chi^2$ due to blinding serves as our principle metric for successful blinding, as it measures the extent to which blinding preserves the internal consistency of the individual observable vector components.

For the fiducial test, the impact of blinding on the goodness of fit was below the expected rms statistical fluctuations in χ^2 for all realizations (Fig. 5), and the change in best-fitting parameters was generally well predicted by the input blinding shift (Fig. 7). As expected from the leading-order analysis for non-linear models (Section 3.3), the $\Delta\chi^2$ figure of merit for blinding scales as a power of the product of the sizes of the blinding shift $\Delta\Theta$ and the 'truth shift' $\Theta_{\rm obs} - \Theta_{\rm ref}$. We also verified that those results are not significantly affected when Gaussian noise is added to the simulated data (Appendix C). We noted one potential cause for concern, in that for a small number of realizations with large truth offset $\Theta_{\rm obs} - \Theta_{\rm ref}$, the blinding transformation resulted in large changes to the best-fitting values of nominally unblinded parameters, which were large enough to push the posterior into a prior boundary.

When we stress the blinding transformation with the nuisance-parameter test case, the typical $\Delta\chi^2$ performance worsened somewhat (Figs 8–10). The majority of realizations in this latter test still fall below the criterion $\Delta\chi^2 < 30$ (i.e. less than the standard deviation of χ^2), but 4 out of 100 realizations exceeded that value, indicating a poor fit of our model to the blinded data. These realizations are found, however, to have input parameters $\Theta_{\rm obs}$ that are very far from the reference parameters $\Theta_{\rm ref}$ assumed for blinding, both in the sense that they would have low probability under the priors on nuisance parameters, and that they are the equivalent of $\approx 50\sigma$ unlikely for the Y3-3 × 2pt parameter errors at $\Theta_{\rm ref}$. An appropriate choice of $\Theta_{\rm ref}$ will preclude the appearance of these high-$\Delta\chi^2$ cases in the real Y3-3 × 2pt analysis.

While we have demonstrated that equation (7) defines a viable blinding scheme for Y3-3 × 2pt wCDM analyses, it is possible that other experiments will encounter cases where the non-linearities in the data model generate unacceptably large $\Delta\chi^2$ over the ranges of $\Delta\Theta$ and $\Theta_{\rm obs}$, which are necessary for effective blinding over the full parameter space allowed by priors. This issue is not necessarily prohibitive, however: one can imagine additional steps in the blinding procedure to account for it (and which could potentially also be used to handle cases where prior-boundary collisions occur even with acceptably small $\Delta\chi^2$). For example, if all pipeline checks pass on blinded data, but the resultant $\Theta_{\rm bl}$ encroaches on the boundaries of priors on nuisance parameters, a designated member of the collaboration could look at how the posterior changes when the data are blinded using a different randomly drawn blinding shift $\Delta\Theta$. Alternatively, the person tasked with generating the blinding shifts could generate two or more $f^{\rm (add)}$ shifts that use $\Theta_{\rm ref}$ values in distinct parts of the parameter space. The collaboration's criterion could be that the data are accepted if any one of these blinded data sets generate an acceptable χ^2. This would allow valid data to pass when its $\Theta_{\rm obs}$ is sufficiently close to any one of the $\Theta_{\rm ref}$s,

while truly systematic errors would still be likely to fail the $\Delta\chi^2$ criterion. Other strategies may be viable as well.

There are a number of considerations one should take into account when deciding whether and how to adopt the summary-statistic blinding transformation described in this paper. Summary-statistic blinding has the advantage that there is relatively low overhead for implementation: it can make use of existing analysis infrastructure, since the blinding factors are computed using the same theory prediction machinery needed for parameter estimation. To make the blinding more robust, it should ideally be implemented as an automatic step when summary statistics are measured from the data. It is a trivial matter for a collaboration member to infer the blinding shifts and unblind their data, if the blinding code is freely available: All one needs to do is run a zero vector through the blinding subroutine. Thus some level of self-control and trust are still required for successful blinding – we are not proposing a foolproof cryptographic system.

It is also worth considering whether multiple stages of blinding should be adopted: Especially for a new blinding technique and new analyses, having a step of parameter-level blinding (hiding numbers on parameter constraint contours) even after unblinding the summary statistics, can be useful. It is also worth considering how to check or protect against spurious cases where this transformation may lead to undesired behaviour, like pushing an acceptable unblinded parameter values past its bound from the prior. It is also important to keep in mind that blinding necessarily adds time to an analysis and in particular that using this transformation will require MCMC chains for parameter estimation to be re-run when it is time to unblind. One can argue that this serves as a feature rather than a bug: The barrier to unblinding can help force a collaboration of busy people with divided attention to pause and consider the status of an analysis before proceeding.

This summary-statistic blinding transformation is implemented as part of the ongoing DES Y3-3 $\times$ 2pt analysis. In practice, the blinding transformation is applied using a script that runs automatically when the 2PCF are measured from galaxy catalogues. This script uses a string seed to pseudo-randomly draw a blinding shift in parameter space, which uses the same configuration files as the parameter estimation pipeline to compute and apply blinding factors to the measured 2PCF. The same transformation will also be applied to the combined analysis of the 3 $\times$ 2pt data with CMB lensing. Looking further forward, summary statistic blinding has the potential for broad applicability to many kinds of multiprobe cosmological analyses. It would be interesting to study its applicability to summary statistics for observables beyond 2PCF, such as supernovae, galaxy cluster number counts, or spectroscopic galaxy clustering measurements.

Blinding to protect results against human bias is essential in modern observational cosmology, where complex analyses combining data from multiple observables are leveraged to make increasingly precise constraints. Whereas powerful blinding techniques had already been devised in experimental particle physics, they do not naturally translate to cosmology, particularly to multiprobe analyses that can generally not be separated into distinct 'signal' and 'background' domains. The blinding transformation described in this paper provides a new and promising method for blinding such analyses, which we have demonstrated is applicable to current multiprobe analyses like those being done for DES. An important property of this blinding transformation is it becomes more effective (lower $\Delta\chi^2$, see Section 3.3) as experiments evolve to higher precision and our priors and prejudices focus on smaller regions of parameter space, and non-linear components of the model shrink in comparison to the linear. This makes it a promising potential tool for future cosmological analysis. Of course, one should explicitly investigate how the performance of summary statistic blinding changes as noise on the data decreases to levels like one might expect for future surveys like DESI, LSST, Euclid, and WFIRST.

ACKNOWLEDGEMENTS

This paper has gone through internal review by the DES collaboration.

JM has been supported by the Porat Fellowship at Stanford University and by the Rackham Graduate School at the University of Michigan through a Predoctoral Fellowship. GB has been supported by grants AST-1615555 from the US National Science Foundation, and DE-SC0007901 from the US Department of Energy. DH has been supported by DOE under Contract DE-FG02-95ER40899 and NSF under contract AST-1813834.

The analysis and framing of this paper benefited from discussions at the 'Blind Analysis in High Stakes Survey Science: When, Why, and How?' workshop[12] held in 2017 March at KIPAC/SLAC. We thank the organizers and attendees of that workshop for sharing their insight and experiences.

The analysis made use of the software tools scipy (Jones et al. 2001), numpy (Oliphant 2006), matplotlib (Hunter 2007), getdist (Lewis 2019), MULTINEST (Feroz & Hobson 2008; Feroz et al. 2009, 2013), cosmosis (Zuntz et al. 2015), and cosmolike (Krause & Eifler 2017). It was supported in part through computational resources and services provided by Advanced Research Computing at the National Science Foundation, Ann Arbor; the National Energy Research Scientific Computing Center (NERSC), a US Department of Energy Office of Science User Facility operated under Contract No. DE-AC02-05CH11231; and the Sherlock cluster, supported by Stanford University and the Stanford Research Computing Center. We would like to thank all of these facilities for providing computational resources and support that contributed to these research results.

Funding for the DES Projects has been provided by the US Department of Energy, the US National Science Foundation, the Ministry of Science and Education of Spain, the Science and Technology Facilities Council of the United Kingdom, the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign, the Kavli Institute for Cosmological Physics at the University of Chicago, Financiadora de Estudos e Projetos, Fundacao Carlos Chagas Filho de Amparo a Pesquisa do Estado do Rio de Janeiro , Conselho Nacional de Desenvolvimento Cientifico e Tecnologico and the Ministerio da Ciencia e Tecnologia, and the Collaborating Institutions in the Dark Energy Survey.

The Collaborating Institutions are Argonne National Laboratories, the University of Cambridge, Centro de Investigaciones Energeticas, Medioambientales y Tecnologicas-Madrid, the University of Chicago, University College London, DES-Brazil, Fermilab, the University of Edinburgh, the University of Illinois at Urbana-Champaign, the Institut de Ciencies de l'Espai (IEEC/CSIC), the Institut de Fisica d'Altes Energies, the Lawrence Berkeley National Laboratory, the University of Michigan, the National Optical Astronomy Observatory, the Ohio State University, the University of Pennsylvania, the University of Portsmouth, and the University of Sussex.

[12]kipac.github.io/Blinding.

REFERENCES

Abbott T. et al., 2018, Phys. Rev. D, 98, 043526
Abbott T. et al., 2019, ApJ, 872, L30
Arisaka K. et al., 1993, Phys. Rev. Lett., 70, 1049
Betoule M. et al., 2014, A&A, 568, A22
Blake C. et al., 2016, MNRAS, 456, 2806
Bridle S., King L., 2007, New J. Phys., 9, 444
Conley A. J. et al., 2006, ApJ, 644, 1
Drlica-Wagner A . et al., 2018, ApJS, 235, 33
Eifler T., Schneider P., Hartlap J., 2009, A&A, 502, 721
Elvin-Poole J . et al., 2018, Phys. Rev. D, 98, 042006
Feroz F., Hobson M. P., 2008, MNRAS, 384, 449
Feroz F., Hobson M. P., Bridges M., 2009, MNRAS, 398, 1601
Feroz F., Hobson M. P., Cameron E., Pettitt A. N., 2013, Open J. Astrophys., 2, 10
Hamana T. et al., 2020, PASJ, 72, 1
Handley W., Lemos P., 2019, Phys. Rev. D, 100, 043504
Heymans C. et al., 2012, MNRAS, 427, 146
Hildebrandt H. et al., 2017, MNRAS, 465, 1454
Howlett C., Lewis A., Hall A., Challinor A., 2012, J. Cosmol. Astropart. Phys., 1204, 027
Hunter J. D., 2007, Comput. Sci. Eng., 9, 90
Joachimi B., Schneider P., Eifler T., 2007, A&A, 477, 43,
Jones E . et al., 2001, SciPy: Open Source Scientific Tools for Python. Available at: http://www.scipy.org/
Kilbinger M. et al., 2009, A&A, 497, 677
Klein J. R., Roodman A., 2005, Annu. Rev. Nucl. Part. Sci., 55, 141
Kowalski M. et al., 2008, ApJ, 686, 749
Krause E., Eifler T., 2017, MNRAS, 470, 2100
Krause E . et al., 2017, Phys. Rev. D, preprint (arXiv:1706.09359)
Kuijken K. et al., 2015, MNRAS, 454, 3500
Lewis A., 2019, GetDist: MCMC Sample Analysis, Plotting and GUI.Available at: https://github.com/cmbant/getdist
Lewis A., Challinor A., Lasenby A., 2000, ApJ, 538, 473
Nelder J. A., Mead R., 1965, Comput. J., 7, 308
Oliphant T. E., 2006, A Guide to NumPy. Trelgol Publishing.Available at: https://www.scipy.org/citing.html
Rubin D. et al., 2015, ApJ, 813, 137
Sellentin E., 2020, MNRAS, 492, 3396
Suyu S. H. et al., 2013, ApJ, 766, 70
Suyu S. H. et al., 2017, MNRAS, 468, 2590
Suzuki N. et al., 2012, ApJ, 746, 85
Takahashi R., Sato M., Nishimichi T., Taruya A., Oguri M., 2012, ApJ, 761, 152
Troxel M. A . et al., 2018, Phys. Rev. D, 98, 043528
von der Linden A. et al., 2014, MNRAS, 439, 2
Zhang B. R., Childress M. J., Davis T. M., Karpenka N. V., Lidman C., Schmidt B. P., Smith M., 2017, MNRAS, 471, 2254
Zuntz J . et al., 2015, Astron. Comput., 12, 45
Zuntz J . et al., 2018, MNRAS, 481, 1149

APPENDIX A: LEADING-ORDER BEHAVIOUR OF BLINDED OBSERVABLE VECTORS

We can safely assume that the model $\boldsymbol{d}(\Theta)$ is analytic and can be expanded about the reference parameters $\Theta_{\rm ref}$ in a Taylor series. We can, without loss of generality, set $\Theta_{\rm ref} = 0$ and $\boldsymbol{d}(\Theta_{\rm ref}) = 0$ in this section. The Taylor expansion becomes

$$d_i(\Theta) = d_{i,\alpha}\theta_\alpha + \frac{1}{2}d_{i,\alpha\beta}\theta_\alpha\theta_\beta + O(\theta^3), \tag{A1}$$

where we adopt the usual conventions that repeated indices within a term indicate summation, and indices after the comma denote derivatives taken at the reference parameters. For clarity, we will use Latin indices to indicate dimensions in data space, and Greek indices for parameter-space dimensions. We will also, in this section, define

$$\boldsymbol{s} = \Delta\Theta$$
$$\boldsymbol{t} = \Theta_{\rm obs} - \Theta_{\rm ref},$$

the latter being the true cosmology for the observed Universe. In the noise-free case, we can apply the quadratic approximation in (A1) to the blinding equation (7) to obtain the blinded observable vector

$$\hat{d}_i^{\rm bl} = d_{i,\alpha}(s_\alpha + t_\alpha) + \frac{1}{2}d_{,i,\alpha\beta}\,(s_\alpha s_\beta + t_\alpha t_\beta). \tag{A2}$$

We seek the best-fitting parameters $\Theta^{\rm bl}$ by minimizing the χ^2 of the solution

$$\chi^2(\Theta^{\rm bl}) = \left[\hat{d}_i^{\rm bl} - d_i(\Theta^{\rm bl})\right] F_{ij} \left[\hat{d}_j^{\rm bl} - d_j(\Theta^{\rm bl})\right], \tag{A3}$$

where we define $F = C_{\boldsymbol{d}}^{-1}$ to be the inverse of the observational covariance matrix. We take F to be independent of the model parameters. In the linear limit, it is easy to see that the blinding shift is always exact, $\Theta^{\rm bl} = \boldsymbol{s} + \boldsymbol{t}$, so we introduce the correction term $\tilde{\theta}$ such that

$$\tilde{\theta} = \Theta^{\rm bl} - \boldsymbol{s} - \boldsymbol{t}. \tag{A4}$$

With this definition, we can write the data differential to leading order in each of $\tilde{\theta}$, s, and t as

$$\hat{d}_i^{\rm bl} - d_i(\Theta^{\rm bl}) \approx -d_{i,\alpha}\tilde{\theta}_\alpha - d_{i,\alpha\beta}\left[s_\alpha t_\beta + \tilde{\theta}_\alpha(s_\beta + t_\beta) + \frac{1}{2}\tilde{\theta}_\alpha\tilde{\theta}_\beta\right]. \tag{A5}$$

Upon substituting this Taylor expansion back into (A3), we can find the blinding shift adjustment $\tilde{\theta}$ that yields the minimal χ^2. Again, retaining only leading-order terms in $\tilde{\theta}$, s, and t:

$$\tilde{\theta}_i \approx -(D^T F D)^{-1} D^T C_{\boldsymbol{d}}^{-1}\boldsymbol{q} \tag{A6}$$

$$\begin{aligned}\Delta\chi^2 &\approx (P\boldsymbol{q})^T\, C_{\boldsymbol{d}}^{-1}\,(P\boldsymbol{q}) \\ &\leq \boldsymbol{q}^T C_{\boldsymbol{d}}^{-1}\boldsymbol{q}\end{aligned} \tag{A7}$$

where we have defined the first derivative matrix and the quadratic data perturbation, respectively, as

$$D_{i\alpha} \equiv d_{i,\alpha}$$
$$q_i \equiv d_{i,\alpha\beta}s_\alpha t_\beta.$$

P is a projection matrix that removes the portion of the non-linear data shift $\boldsymbol{q}$, which can be fitted by a shift in parameters:

$$P \equiv I - D(D^T F D)^{-1} D^T C_{\boldsymbol{d}}^{-1}. \tag{A8}$$

From these equations, several properties of the additive blinding transformation are apparent. First, the transformation is exact, in the sense that $\Delta\chi^2 = 0$, in *any* of the following conditions:

(i) The blinding shift $\boldsymbol{s} = \Delta\Theta$ is zero.
(ii) The true cosmology equals the reference cosmology, $\boldsymbol{t} = \Theta_{\rm obs} - \Theta_{\rm ref} = 0$.
(iii) The model is linear, $d_{i,\,\alpha\beta} = 0$.
(iv) The derivative matrix is invertible, in which case the projector matrix $P = 0$, because any point in data space can be fit exactly with proper choice of parameters, at least locally.

Secondly, we see that a quadratic term in the data model leads to a deviation $\tilde{\theta}$ between the naive blinded cosmology parameters estimate $\Theta^{\rm bl} = \Theta_{\rm obs} + \Delta\Theta$, which scales as the product of the blinding shift $\boldsymbol{s}$ and the 'truth shift' $\boldsymbol{t}$.

Thirdly, the $\Delta\chi^2$ in a quadratic approximation to the data model will scale as the product of the squares of the two shifts, and inversely with the measurement covariance matrix C_d, $\Delta\chi^2 \propto s^2 t^2 / C_d$. The constant of proportionality will depend on the relations between the directions of the blinding shift, the truth shift, the curvature of the model, and the covariance matrix of the observations. If we include terms beyond quadratic in the data model, we will find the dependence of $\Delta\chi^2$ is of even higher order in st.

APPENDIX B: MODELLING FOR 3X2PT OBSERVABLE VECTOR

The theory predictions for the 3 × 2pt observable vectors are computed as follows. First, the non-linear matter power spectrum $P(k, z)$ is computed using a combination of CAMB (Lewis, Challinor & Lasenby 2000; Howlett et al. 2012) and HALOFIT (Takahashi et al. 2012). Then, using the Limber approximation, we integrate to obtain the angular power spectra between the sets of tracers we are studying. For the correlation between the ith redshift bin of tracer A and the jth bin of tracer B, the angular power spectrum is

$$C^{ij}_{AB}(\ell) = \int \mathrm{d}z \frac{H(z)}{c\,\chi^2(z)} W^i_A(z) W^j_B(z) P(k,z)\Big|_{k=(\ell+\frac{1}{2})/\chi(z)}, \quad \text{(B1)}$$

where χ is the comoving radial distance, and the weight functions for galaxy number density g and weak lensing convergence κ are defined as

$$W^i_g(z,k) = n_i(z)\, b_i, \quad \text{(B2)}$$

$$W^i_\kappa(z) = \left(\frac{3H_0^2\Omega_\mathrm{m}}{2c}\right)\left(\frac{\chi(z)}{a(z)\,H(z)}\right) \times \int_z^\infty \mathrm{d}z' n_i(z') \frac{\chi(z')-\chi(z)}{\chi(z')}. \quad \text{(B3)}$$

Here $n_i(z)$ is the normalized redshift distribution and b_i is the galaxy bias of galaxies in bin i. We then perform Fourier transformations to convert these angular spectra into real-space angular correlation functions, which can be compared to data. The galaxy–galaxy correlation is

$$w^{ij}(\theta) = \sum_\ell \frac{2\ell+1}{4\pi} P_\ell(\cos\theta)\, C^{ij}_{gg}(\ell), \quad \text{(B4)}$$

where $P_\ell(x)$ is a Legendre polynomial of order ℓ. In the flat-sky approximation, where sums over spherical harmonics are converted to two-dimensional Fourier modes, the predicted angular correlations between the shears of galaxies in tomographic bins i and j are

$$\xi^{ij}_+(\theta) = \int \frac{\mathrm{d}\ell\,\ell}{2\pi} J_0(\ell\theta)\, C^{ij}_{\kappa\kappa}(\ell), \quad \text{(B5)}$$

$$\xi^{ij}_-(\theta) = \int \frac{\mathrm{d}\ell\,\ell}{2\pi} J_4(\ell\theta)\, C^{ij}_{\kappa\kappa}(\ell), \quad \text{(B6)}$$

$$\gamma^{ij}_t(\theta) = \int \frac{\mathrm{d}\ell\,\ell}{2\pi} J_2(\ell\theta)\, C^{ij}_{g\kappa}(\ell), \quad \text{(B7)}$$

where $J_m(x)$ is a Bessel function of the first kind of order m. In practice, these calculations are done using the function `tp-stat_via_hankel`: from the NICAEA software[13] (Kilbinger et al. 2009).

Nuisance parameters are included as follows. The photo-z bias parameters Δz^x_i, where x = source or lens, have the effect of shifting the redshift distributions of the samples of galaxies:

$$n^x_i(z) \quad \rightarrow \quad n^x_i\left(z - \Delta z^x_i\right). \quad \text{(B8)}$$

Shear calibration parameters m are defined so the measured shear for a galaxy is $\gamma_\mathrm{meas} = (1+m)\gamma_\mathrm{true}$. They modify the 2PCF involving source galaxies via

$$\xi^{ij}_\pm(\theta) \rightarrow (1+m_i)(1+m_j)\xi^{ij}_\pm(\theta), \quad \text{and} \quad \text{(B9)}$$

$$\gamma^{ij}_t(\theta) \rightarrow (1+m_j)\gamma^{ij}_t(\theta). \quad \text{(B10)}$$

The linear intrinsic alignment model used in our analysis modifies the shear convergence weight function (equation B3) via

$$W^i_\kappa(z) \rightarrow W^i_\kappa(z) - \left[A_\mathrm{IA}\left(\frac{1+z}{1+z_0}\right)^{\alpha_\mathrm{IA}} \frac{C_1\rho_{m0}}{D(z)}\right]\frac{\mathrm{d}n^i}{\mathrm{d}z}, \quad \text{(B11)}$$

where $C_1 = 0.0134/\rho_\mathrm{crit}$ is a normalization constant calibrated based on previous observations (Bridle & King 2007).

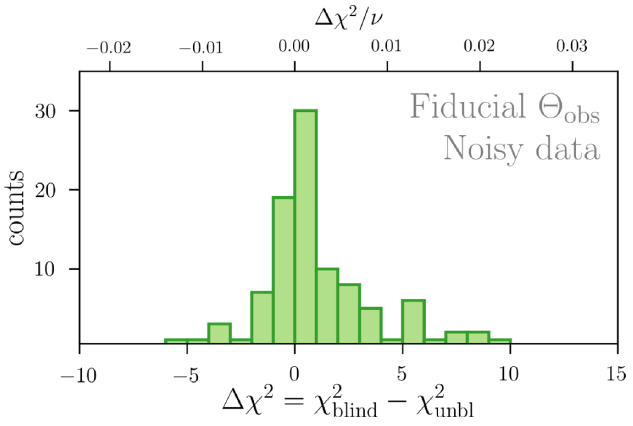

Figure C1. Same as Fig. 5, but with Gaussian noise added to the fiducial ensemble of unblinded observable vectors: histogram of $\Delta\chi^2$ values.

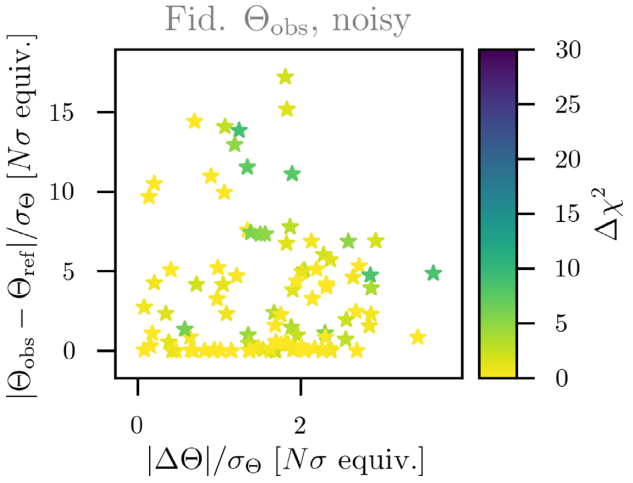

Figure C2. Same as Fig. 6, but with Gaussian noise added to the ensemble of unblinded observable vectors: dependence of $\Delta\chi^2$ on the parameter-space distances associated with $\Theta_\mathrm{obs} - \Theta_\mathrm{ref}$ and $\Delta\Theta$.

[13] www.cosmostat.org/software/nicaea

APPENDIX C: ADDITIONAL RESULTS FOR NOISY DATA

We performed a variation of our fiducial test with noisy measured observable vectors. Here we used the same true cosmology $\Theta_{\rm obs}$ and blinding shifts $\Delta\Theta$ as in Section 4.5. The only difference is that after computing the theory prediction at $\Theta_{\rm obs}$ to generate a synthetic measured observable vector, we added a realization of Gaussian noise produced using a Cholesky decomposition of the covariance C_d.

Results for the noise-added version of our fiducial test are shown here. Fig. C1 shows a histogram of the resulting $\Delta\chi^2$ values. Fig. C2 shows how $\Delta\chi^2$ depends on the magnitude of the distances in parameter space associated with the blinding shift $\Delta\Theta$ and the difference between the true cosmology and that assumed for blinding $\Theta_{\rm obs} - \Theta_{\rm ref}$. There is more scatter in the relations, but, otherwise, the results for this test are not substantially different from those shown for the noiseless test presented in Section 4.5. We additionally confirmed for a few realizations that adding noise to the observable vector does not significantly change how blinding affects posterior contours like those shown in FIG. 4.

[1]*Kavli Institute for Particle Astrophysics & Cosmology, PO Box 2450, Stanford University, Stanford, CA 94305, USA*
[2]*Department of Physics, University of Michigan, Ann Arbor, MI 48109, USA*
[3]*Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA*
[4]*Department of Physics & Astronomy, University College London, Gower Street, London WC1E 6BT, UK*
[5]*Max Planck Institute for Astrophysics, Karl-Schwarzschild-Strasse 1, D-85748 Garching, Germany*
[6]*Department of Astronomy/Steward Observatory, University of Arizona, 933 North Cherry Avenue, Tucson, AZ 85721-0065, USA*
[7]*SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA*
[8]*Fermi National Accelerator Laboratory, PO Box 500, Batavia, IL 60510, USA*
[9]*Instituto de Fisica Teorica UAM/CSIC, Universidad Autonoma de Madrid, E-28049 Madrid, Spain*
[10]*LSST, 933 North Cherry Avenue, Tucson, AZ 85721, USA*
[11]*Physics Department, 2320 Chamberlin Hall, University of Wisconsin-Madison, 1150 University Avenue, Madison, WI 53706-1390, USA*
[12]*CNRS, UMR 7095, Institut d'Astrophysique de Paris, F-75014 Paris, France*
[13]*Sorbonne Universités, UPMC Univ Paris 06, UMR 7095, Institut d'Astrophysique de Paris, F-75014 Paris, France*
[14]*Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), E-28040 Madrid, Spain*
[15]*Laboratório Interinstitucional de e-Astronomia – LIneA, Rua Gal. José Cristino 77, Rio de Janeiro, RJ-20921-400, Brazil*
[16]*Department of Astronomy, University of Illinois at Urbana-Champaign, 1002 W. Green Street, Urbana, IL 61801, USA*
[17]*National Center for Supercomputing Applications, 1205 West Clark St., Urbana, IL 61801, USA*
[18]*Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, E-08193 Bellaterra (Barcelona), Spain*
[19]*INAF-Osservatorio Astronomico di Trieste, via G. B. Tiepolo 11, I-34143 Trieste, Italy*
[20]*Institute for Fundamental Physics of the Universe, Via Beirut 2, I-34014 Trieste, Italy*
[21]*Observatório Nacional, Rua Gal. José Cristino 77, Rio de Janeiro, RJ-20921-400, Brazil*
[22]*Department of Physics, IIT Hyderabad, Kandi, Telangana 502285, India*
[23]*Excellence Cluster Origins, Boltzmannstr. 2, D-85748 Garching, Germany*
[24]*Faculty of Physics, Ludwig-Maximilians-Universität, Scheinerstr. 1, D-81679 Munich, Germany*
[25]*Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Dr., Pasadena, CA 91109, USA*
[26]*Santa Cruz Institute for Particle Physics, Santa Cruz, CA 95064, USA*
[27]*Institut d'Estudis Espacials de Catalunya (IEEC), E-08034 Barcelona, Spain*
[28]*Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans, s/n, E-08193 Barcelona, Spain*
[29]*Kavli Institute for Cosmological Physics, University of Chicago, Chicago, IL 60637, USA*
[30]*Department of Astronomy, University of Michigan, Ann Arbor, MI 48109, USA*
[31]*Department of Physics, Stanford University, 382 Via Pueblo Mall, Stanford, CA 94305, USA*
[32]*Department of Physics, ETH Zurich, Wolfgang-Pauli-Strasse 16, CH-8093 Zurich, Switzerland*
[33]*Center for Astrophysics | Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 02138, USA*
[34]*Australian Astronomical Optics, Macquarie University, North Ryde, NSW 2113, Australia*
[35]*Lowell Observatory, 1400 Mars Hill Rd, Flagstaff, AZ 86001, USA*
[36]*George P. and Cynthia Woods Mitchell Institute for Fundamental Physics and Astronomy, Department of Physics and Astronomy, Texas A&M University, College Station, TX 77843, USA*
[37]*Department of Astrophysical Sciences, Princeton University, Peyton Hall, Princeton, NJ 08544, USA*
[38]*Institució Catalana de Recerca i Estudis Avançats, E-08010 Barcelona, Spain*
[39]*Department of Physics and Astronomy, Pevensey Building, University of Sussex, Brighton BN1 9QH, UK*
[40]*School of Physics and Astronomy, University of Southampton, Southampton SO17 1BJ, UK*
[41]*Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA*
[42]*Institute of Cosmology and Gravitation, University of Portsmouth, Portsmouth PO1 3FX, UK*
[43]*Department of Physics, Duke University, Durham, NC 27708, USA*
[44]*Cerro Tololo Inter-American Observatory, National Optical Astronomy Observatory, Casilla 603, La Serena, Chile*
[45]*Max Planck Institute for Extraterrestrial Physics, Giessenbachstrasse, D-85748 Garching, Germany*
[46]*Universitäts-Sternwarte, Fakultät für Physik, Ludwig-Maximilians Universität München, Scheinerstr. 1, D-81679 München, Germany*
[47]*Institute for Astronomy, University of Edinburgh, Edinburgh EH9 3HJ, UK*

This paper has been typeset from a TEX/LATEX file prepared by the author.