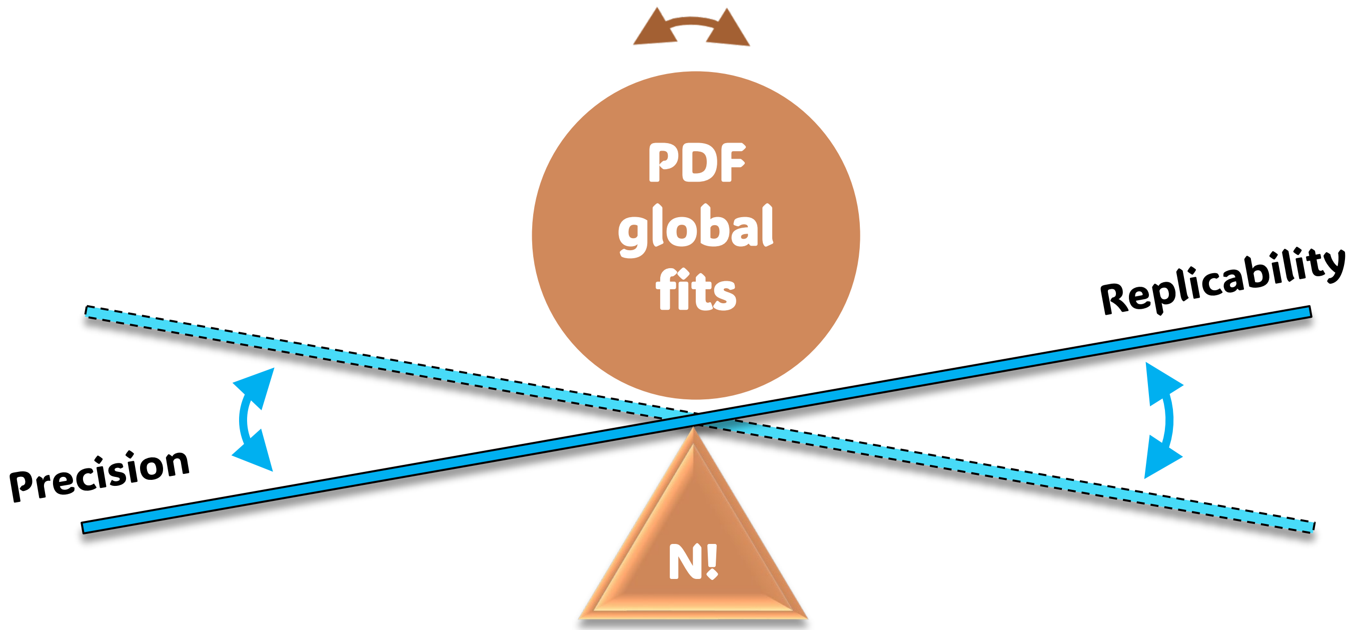
Balancing precision and replicability

A proposal for a PDF4LHC study

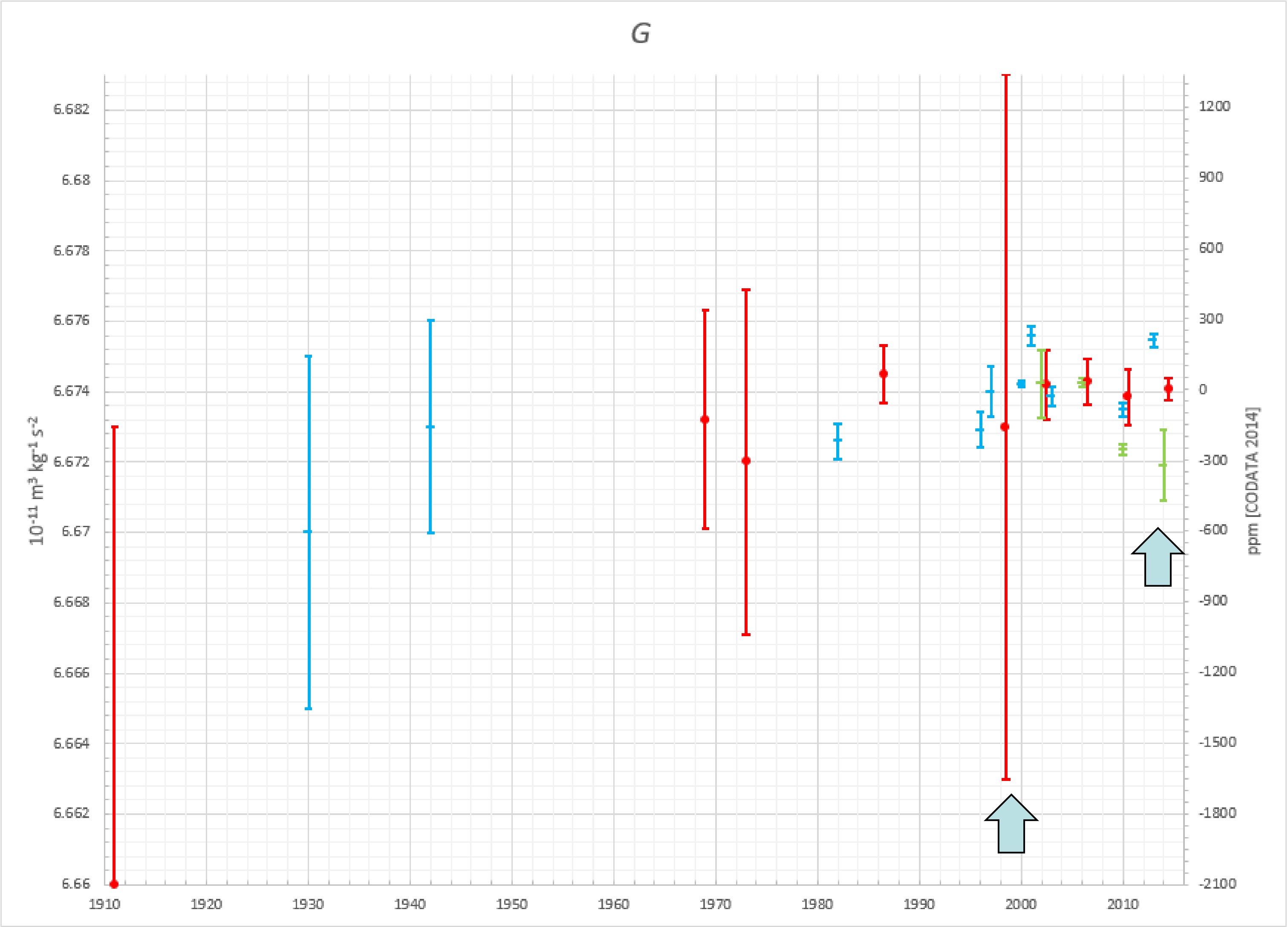
Pavel Nadolsky

Southern Methodist University

With CTEQ-TEA (Tung Et Al.)
Global QCD analysis group



World average for the gravitational constant



Timeline of measurements and recommended values for *G* since 1900: values recommended based on the NIST combination (red), individual torsion balance experiments (blue), other types of experiments (green).

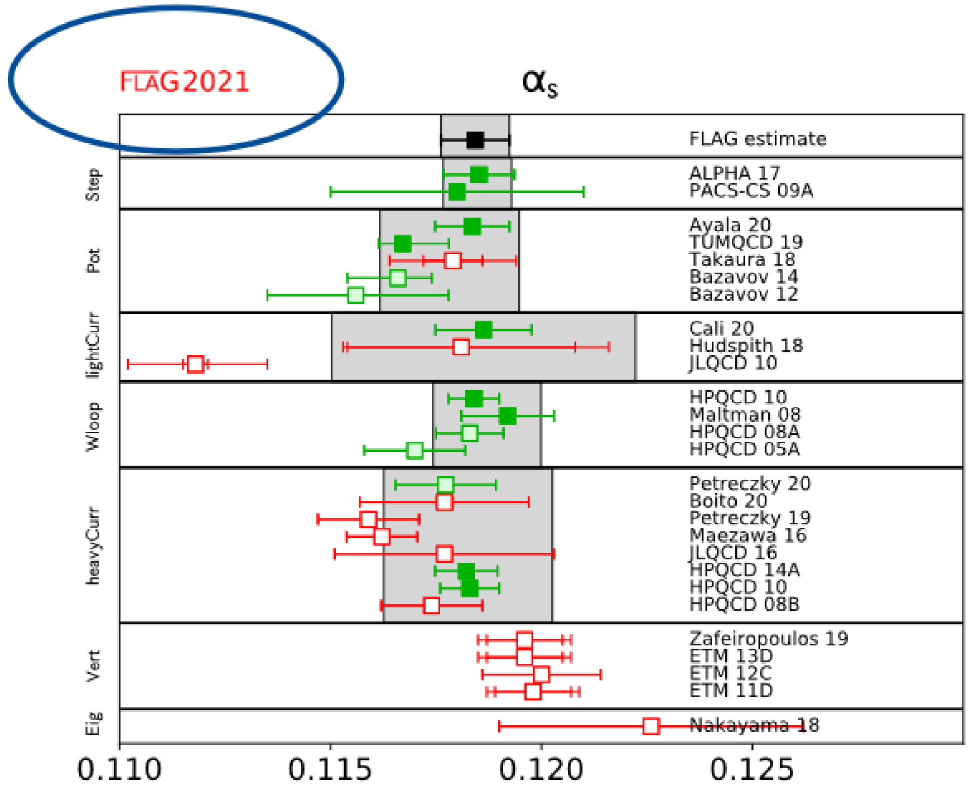
The combination error bars are unstable after 1995

Some precise individual measurements are in a conflict among themselves and post-2014 combination

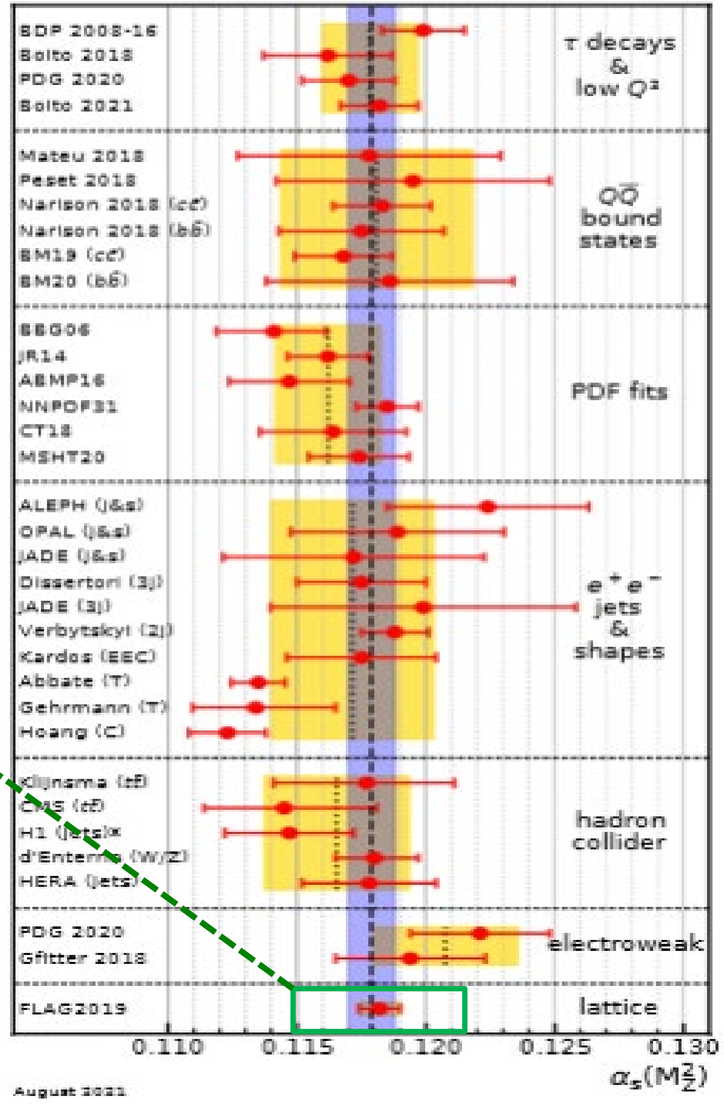
https://en.wikipedia.org/wiki/Gravitational_constant#Modern_value, retrieved on Oct. 22, 2023


The sausage-making of α_s
We (PDG) divide the determinations into 7 categories and take an unweighted fit for each category.
The 6 non-lattice measurements are then averaged with the lattice average provided by the FLAG group
τ decays & low Q²
QQ bound states
DIS & PDF fits
e⁺e⁻ jets & shapes
hadron collider
electroweak
lattice
0.110 0.115 0.120 0.125 0.130
August 2019
Figure 9.2: Summary of determinations of $\alpha_s(M_Z^2)$ from the seven sub-fields discussed in the text. The yellow (light shaded) bands and dotted lines indicate the pre-average values of each sub-field. The dashed line and blue (dark shaded) band represent the final world average value of $\alpha_s(M_Z^2)$.
Les Houches, June 2023

Lattice QCD & world-average α_s combination



arXiv:2111.09849



Lattice determinations of α_s in multiple channels are projected to be [far] more precise than many experiments. Several challenges with combining the eclectic α_s inputs with the current procedure.

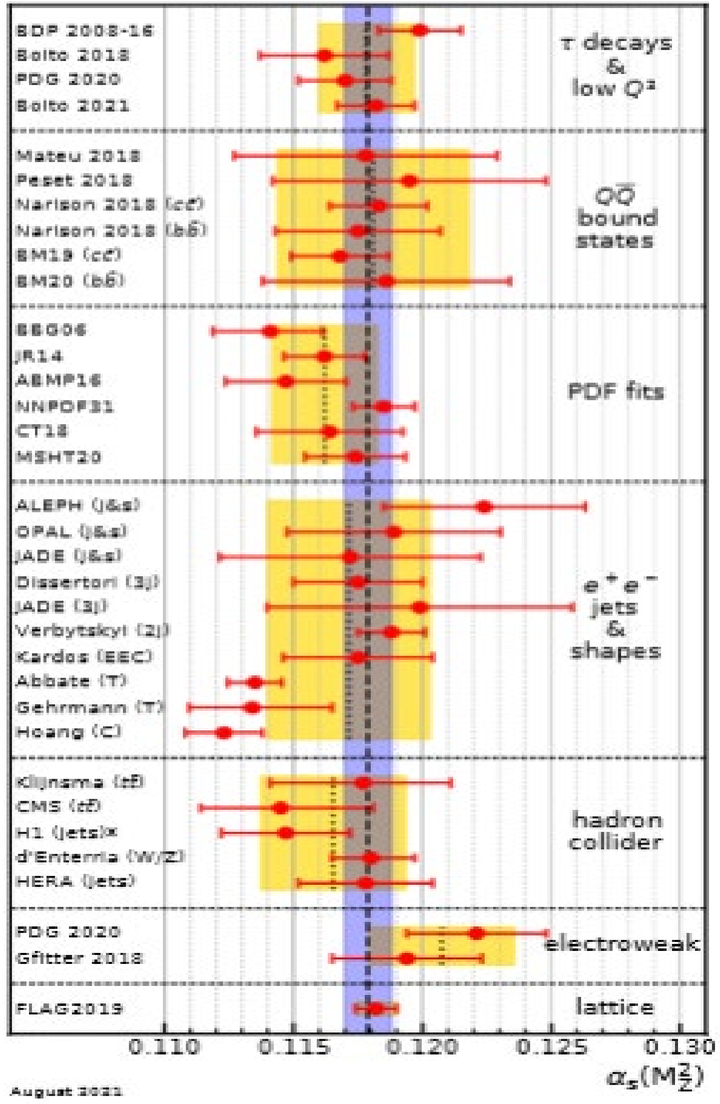
Time to rethink how the world-average α_s combination is performed?

Future measurements of the QCD coupling

individual α_s measurements can reach precision of $\sim 0.1\%$

and symbols: CIPT='contour-improved perturbation theory', FOPT='fixed-order perturbation theory', NP='nonperturbative QCD', SF='structure functions', PS='Monte Carlo parton shower'.

Method	Relative $\alpha_s mZ$ uncertainty Current theory & exp. uncertainties sources	Near (long-term) future theory & experimental progress
(1) Lattice	0.7% Finite lattice spacing & stats. $N^{2,3}$LO pQCD truncation	≈ 0.3% (0.1%) Reduced latt. spacing. Add more observables Add $N^{3,4}$LO, active charm (QED effects) Higher renorm. scale via step-scaling to more observ.
(2) τ decays	1.6% N^3LO CIPT vs. FOPT diffs. Limited τ spectral data	< 1.% Add N^4LO terms. Solve CIPT–FOPT diffs. Improved τ spectral functions at Belle II
(3) $Q\bar{Q}$ bound states	3.3% $N^{2,3}$LO pQCD truncation $m_{c,b}$ uncertainties	≈ 1.5% Add $N^{3,4}$LO & more ($c\bar{c}$), ($b\bar{b}$) bound states Combined $m_{c,b} + \alpha_s$ fits
(4) DIS & PDF fits	1.7% $N^{2,(3)}$LO PDF (SF) fits Span of PDF-based results	≈ 1% (0.2%) N^3LO fits. Add new SF fits: $F_2^{p,d}$, g_i (EIC) Better corr. matrices, sampling of PDF solutions. More PDF data (EIC/LHeC/FCC-eh)
(5) e^+e^- jets & evt shapes	2.6% NNLO+$N^{(1,2,3)}$LL truncation Different NP analytical & PS corrs. Limited datasets w/ old detectors	≈ 1.5% (< 1%) Add $N^{2,3}$LO+N^3LL, power corrections Improved NP corrs. via: NNLL PS, grooming New improved data at B factories (FCC-ee)
(6) Electroweak fits	2.3% N^3LO truncation Small LEP+SLD datasets	(≈ 0.1%) N^4LO, reduced param. uncerts. ($m_{W,Z}$, α, CKM) Add W boson. Tera-Z, Oku-W datasets (FCC-ee)
(7) Hadron colliders	2.4% NNLO(+NNLL) truncation, PDF uncerts. Limited data sets ($t\bar{t}$, W, Z, e-p jets)	≈ 1.5% N^3LO+NNLL (for color-singlets), improved PDFs Add more datasets: Z p_T, p-p jets, σ_i/σ_j ratios,...
World average	0.8%	≈ 0.4% (0.1%)



D. d'Enterria et al., EF QCD, arXiv:2203.08271

Ongoing studies of systematic uncertainties are essential and still insufficient

- from the experiment side

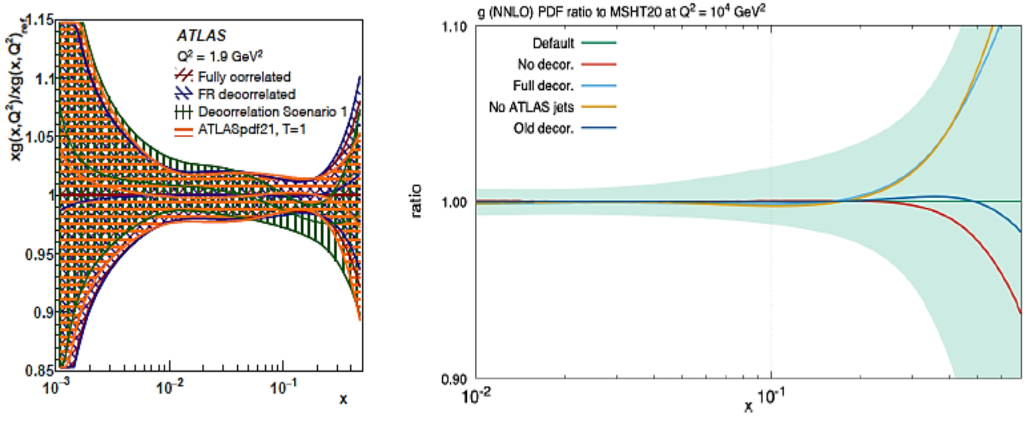


FIG. 9. Difference in the gluon PDF shown in ratio to the ATLASpdf21 (default) gluon(left). This default uses Decorrelation Scenario 2 and this is compared to the use of Full Correlation, Full decorrelation of the flavour response systematic and Decorrelation Scenario 1. The effect of no decorrelation, the default correlation of [9], the decorrelation in [362], and full decorrelation for the MSHT20 gluon (right).

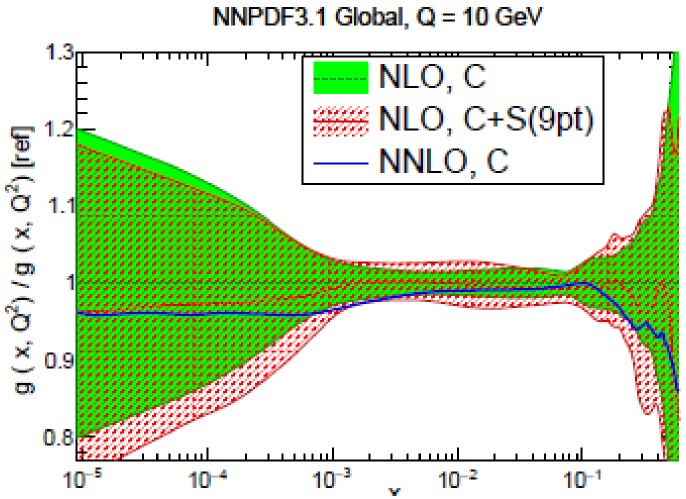
S. Amoroso et al., 2203.13923, Sec. 5.A

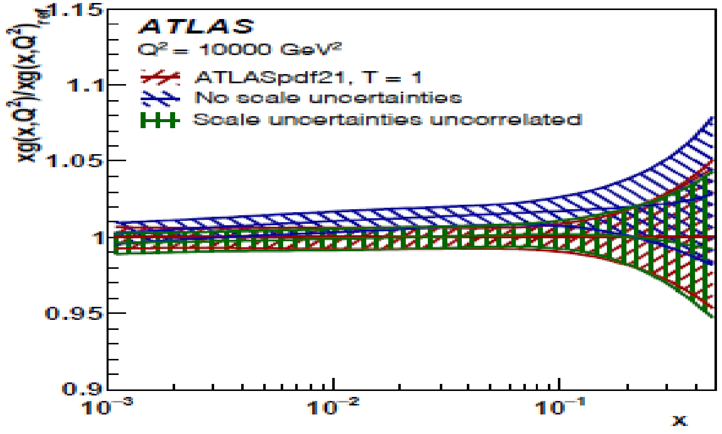
Strong dependence on the definition of corr. syst. errors raises a general concern:

Overreliance on Gaussian distributions and covariance matrices for poorly understood effects may produce very wrong uncertainty estimates

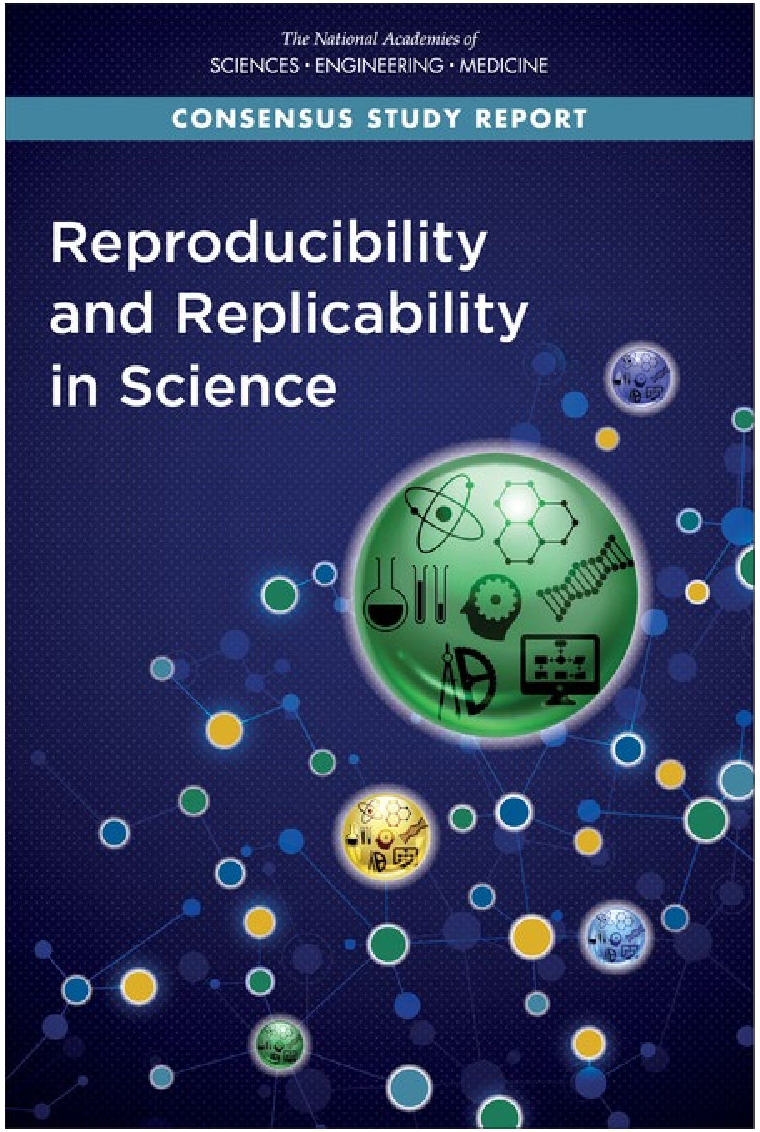
[N. Taleb, Black Swan & Antifragile]

- from the theory side





Examples: studies of theory uncertainties in the PDFs by NNPDF3.1 and ATLAS21



US National Academy of Sciences, Engineering, and Medicine, 2019, https://doi.org/10.17226/25303

Replicability risks for precision QCD

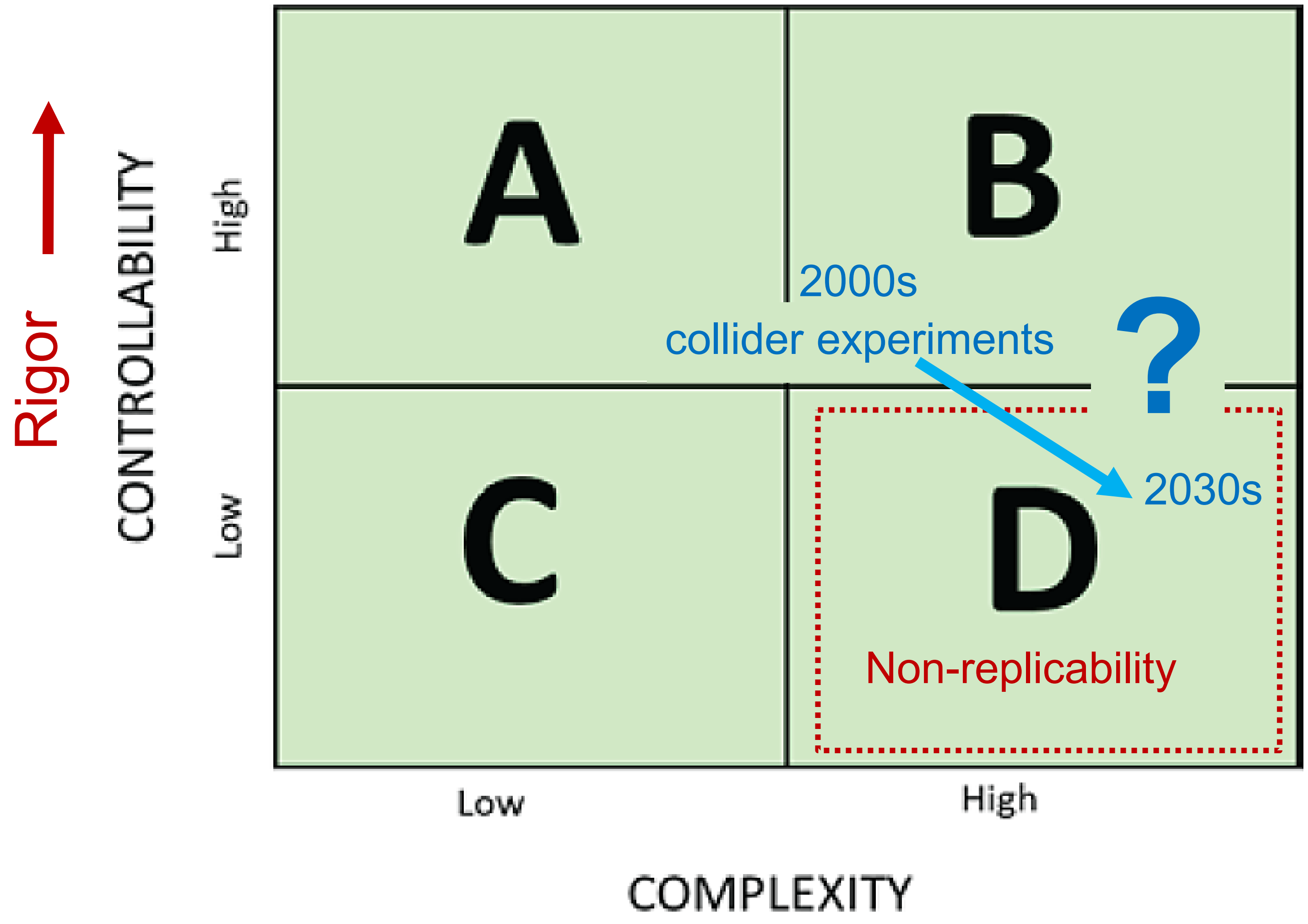
Replicability is a requirement of obtaining consistent results across studies aimed at answering the same scientific question, each of which has its own analysis strategy or data.

Nearly all complex STEM fields encounter replicability challenges.

Modern particle physics is not an exception.

1. It is complex! Is it rigorous enough?
   - Many approaches, especially AI-based ones, increase complexity and are not rigorously understood
2. It often uses wrong prescriptions for estimating epistemic uncertainties
   - Tens to hundreds of systematic uncertainties affect measurements, phenomenology, and lattice QCD

Future scenarios for QCD precision analysis

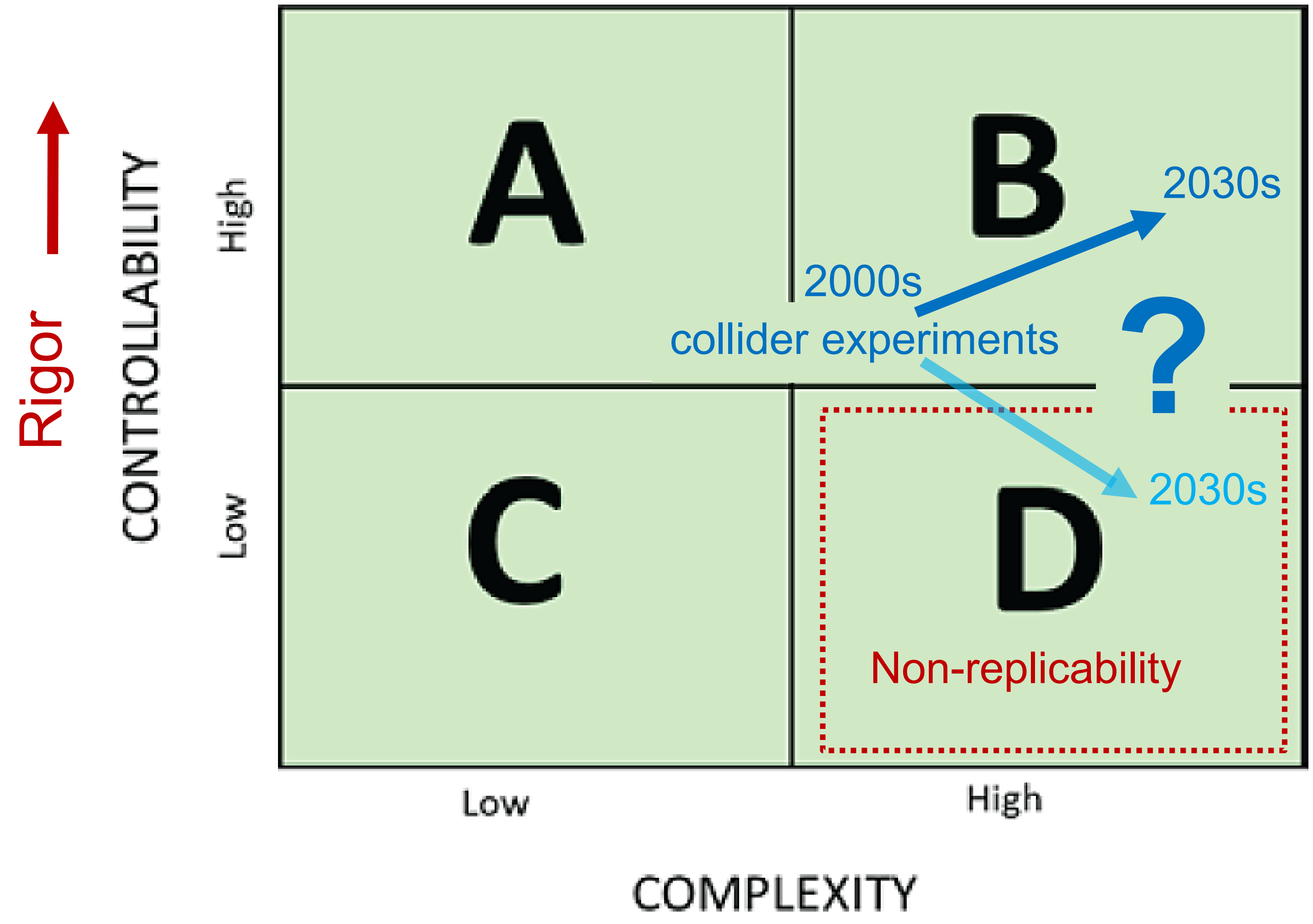


significant risk with current practices

similar to non-replicability issues in other STEM fields

Based on Fig. 5.2 in "REPRODUCIBILITY AND REPLICABILITY IN SCIENCE"

Future scenarios for QCD precision analysis



preferred scenario; requires a coordinated community strategy to adopt the **replicability mindset**

Based on Fig. 5.2 in "REPRODUCIBILITY AND REPLICABILITY IN SCIENCE"

Strategies for improving replicability and reproducibility

Preselection of planned studies based on their likely replicability

Detailed documentation of methods and uncertainty quantification in the publications

Training of researchers in relevant statistical methods

Journal policies that encourage replicability

Support from the funding agencies for the research infrastructure and collaborations focusing on replicability

Support for open publication of the analysis codes and key data, using agreed-upon formats

“Skin-in-the-game” incentives for researchers to produce replicable results

Based on “REPRODUCIBILITY AND REPLICABILITY IN SCIENCE”

Epistemic PDF uncertainty…

…reflects **methodological choices** such as PDF functional forms, NN architecture and hyperparameters, or model for systematic uncertainties

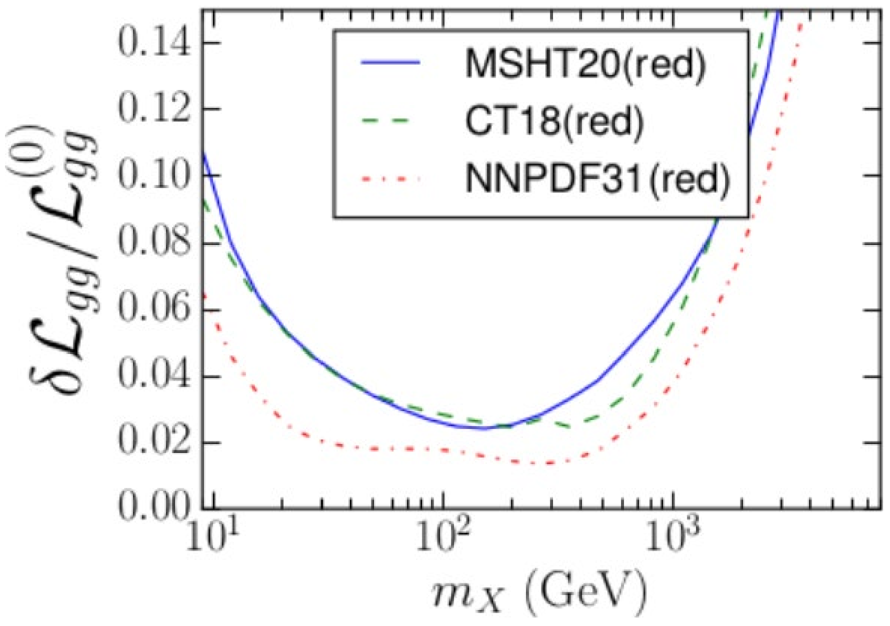
… can dominate the full uncertainty when experimental and theoretical uncertainties are small.

…is associated with the **prior probability**.

… can be estimated by **representative sampling** of the PDF solutions obtained with acceptable methodologies.

⇒ sampling over choices of experiments, PDF/NN functional space, models of correlated uncertainties…

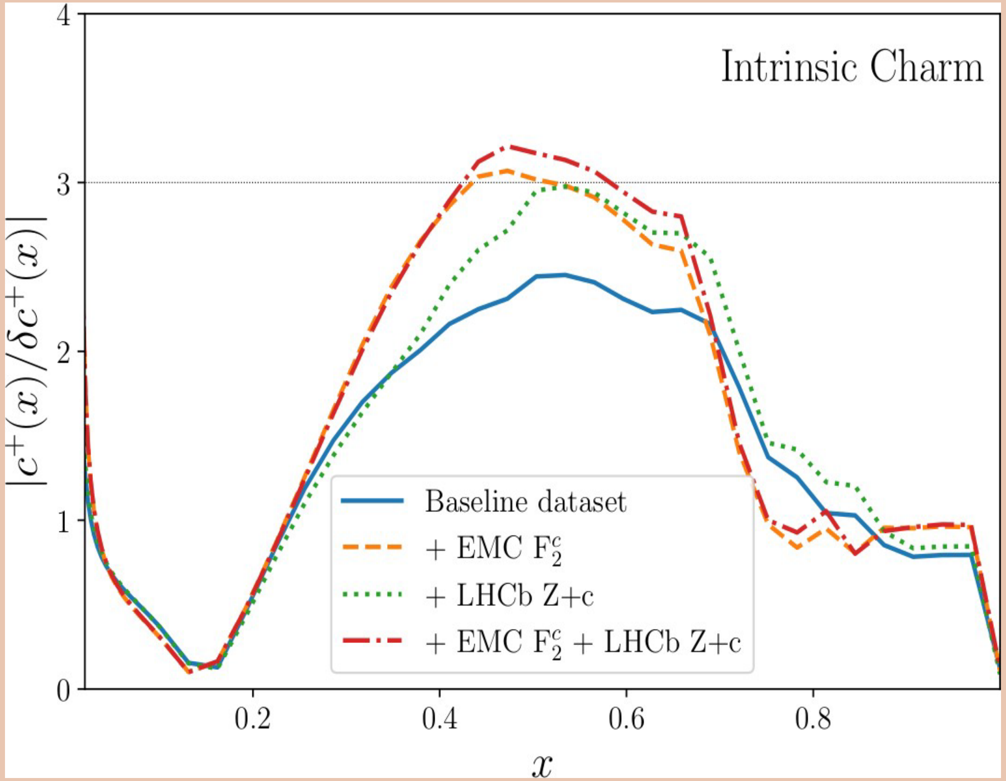
⇒ in addition to sampling over data fluctuations



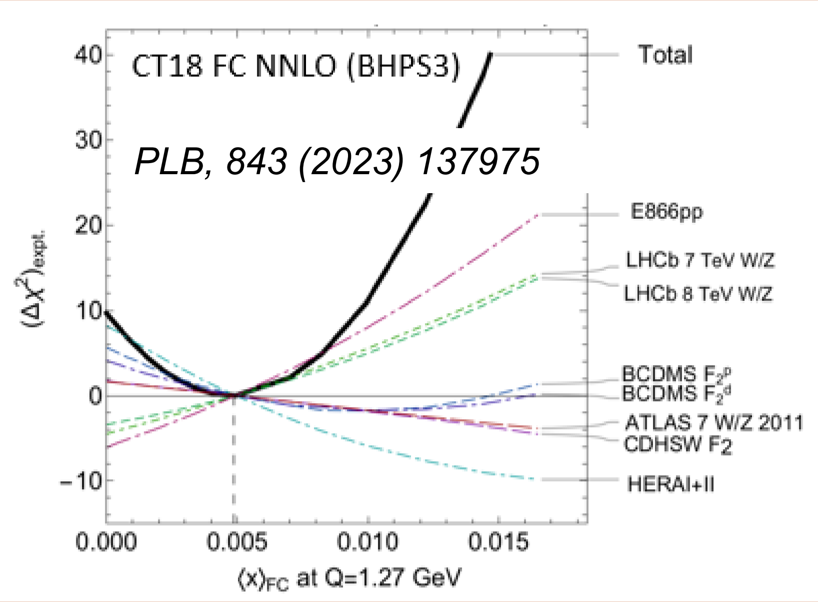
Epistemic uncertainties explain many of the differences among the sizes of PDF uncertainties by CT, MSHT, and NNPDF global fits to the same or similar data

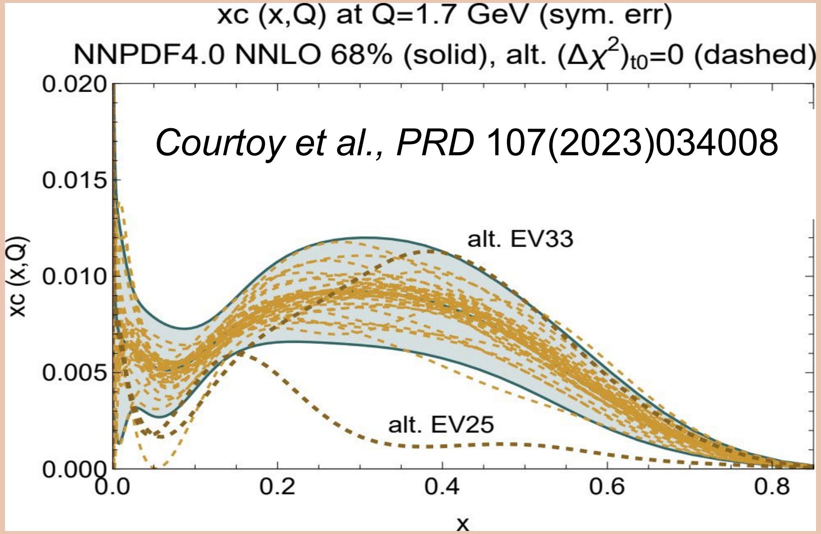
Details in **arXiv:2203.05506**, **arXiv:2205.10444**

How significant is the fitted charm?



NNPDF states a 3σ evidence for $f_{IC}(x, Q_0) \neq 0$ based on the combined constraints from the baseline fit, LHCb $Z + c$ analysis, and EMC F_2^c data





CTEQ-TEA authors find larger uncertainties in each of these sources.

This conclusion is also supported by

- **Lagrange multiplier scans** in the CT18 FC fit [upper figure]
- **hopscotch sampling** of MC replicas in the NNPDF4.0 fitting code [lower figure].

Consequently, $f_{FC}(x, Q_0) \approx 0$ is allowed with high confidence.

Possible contents for the PDF4LHC document

1. Importance of RRR = Reproducibility, Replicability, Rigor
2. Universal factors affecting replicability
3. Recommended practices for improving replicability and reproducibility in the LHC analyses
4. A PDF4LHC wish list for systematic uncertainties
5. “The PDF4LHC challenge”: a scheme to reward PDF fitting groups for quoting PDF uncertainties that are both precise and replicable

Details in backup slides

Specific tips for improving replicability

1. With $O(10-1000)$ free parameters, including nuisance parameters, the $\Delta\chi^2 = 1$ criterion for 1σ PDF uncertainties is almost certainly incomplete. Stop using it “as is”. There are strong mathematical reasons.

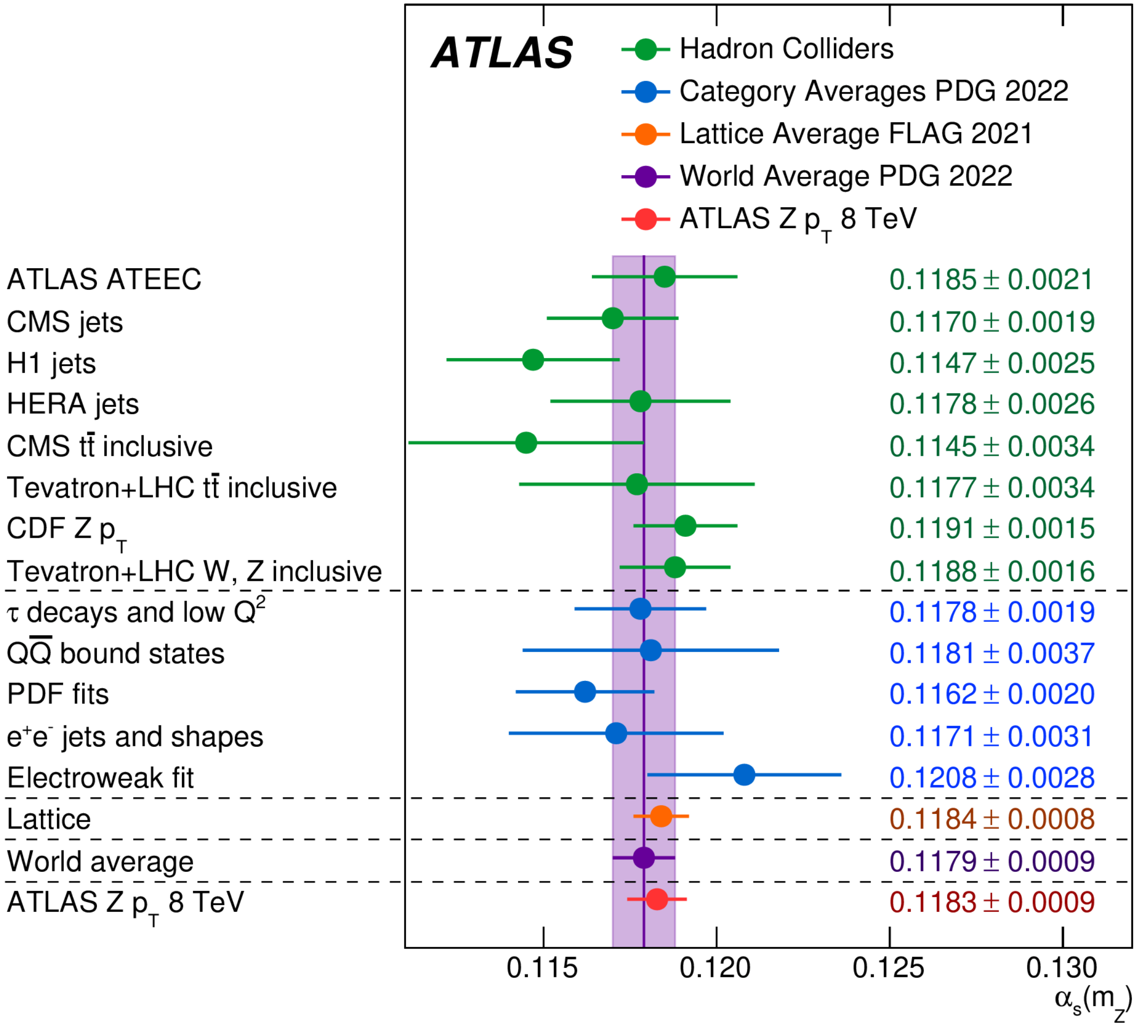
2. Thoroughly estimate the dependence on PDF parametrization forms, NN hyperparameters, and analysis settings when other uncertainties are small.

   - Public tools for this are increasingly available: xFitter, NNPDF code, ePump, Fantômas, MP4LHC,…

ATLAS measures strength of the strong force with record precision

The result showcases the power of the LHC to push the precision frontier and improve our understanding of nature

25 SEPTEMBER, 2023 arXiv:2309.12986



A novel determination of $\alpha_Z(M_Z)$ from $Z\ q_T$ data. However, the PDF uncertainties were not estimated properly

Profiling of global PDFs using $\Delta\chi^2 = 1 \Rightarrow$ Underestimated uncertainties $\Rightarrow$ Non-replicable result

[Details in T.J. Hou et al., 1912.10053, Appendix F]

Two common forms of χ^2 in PDF fits

1. In terms of nuisance parameters $\lambda_{\alpha,exp}$

$$\chi^2 = \sum_{i=1}^{N_{pt}} \frac{\left[D_i + \sum_\alpha \beta_{i,\alpha}^{\rm exp} \lambda_{\alpha,\rm exp} - T_i\right]^2}{s_i^2} + \sum_\alpha \lambda_{\alpha,\rm exp}^2$$

2. In terms of the covariance matrix

algebraic minimization of χ^2 with respect to $\lambda_{\alpha,exp}$

$$\chi^2 = \sum_{i,j}^{N_{pt}} (T_i - D_i)(\mathrm{cov}^{-1})_{ij}(T_j - D_j)$$

$$(\mathrm{cov})_{ij} \equiv s_i^2 \delta_{ij} + \sum_{\alpha=1}^{N_\lambda} \beta_{i,\alpha}\beta_{j,\alpha}, \qquad \beta_{i,\alpha} = \sigma_{i,\alpha} X_i,$$

D_i, T_i, s_i are the central data, theory, uncorrelated error
$\beta_{i,\alpha}$ is the correlation matrix for N_λ nuisance parameters.

Experiments publish $\sigma_{i,\alpha}$ (up to hundreds per data set). To reconstruct $\beta_{i,\alpha}$, we need to decide on the normalizations X_i. Possible choices:

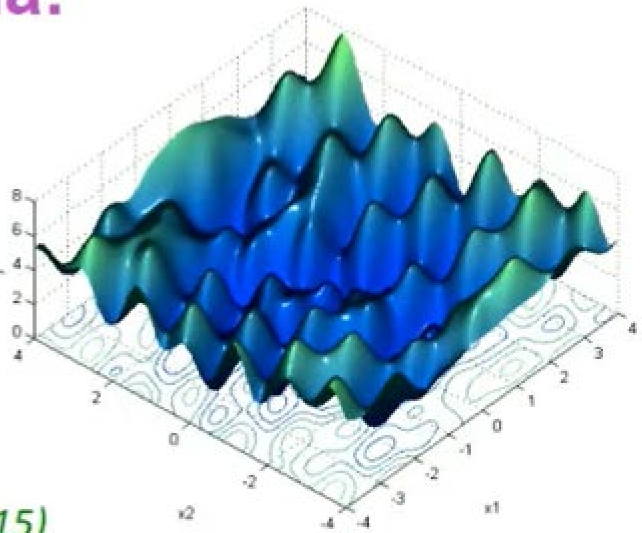
a. $X_i = D_i$: "**exp**erimental scheme"; can result in a bias
b. $X_i = \text{fixed or varied } T_i$: "$\boldsymbol{t_0}$, T, extended T schemes"; can result in (different) biases

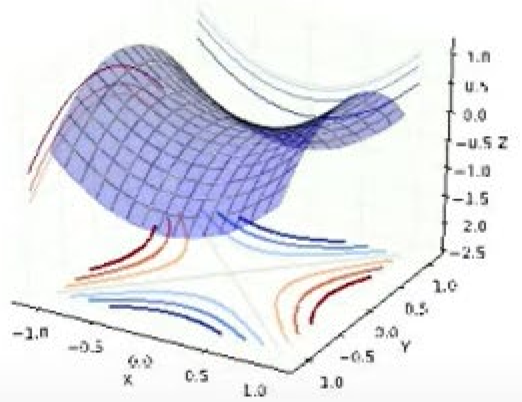
Not so terrible local minima: convexity is not needed

Myth busted:

- Local minima dominate in low-D, but saddle points dominate in high-D
- Most local minima are relatively close to the bottom (global minimum error)

(Dauphin et al NIPS'2014, Choromanska et al AISTATS'2015)





Global minimum: all $\frac{\partial^2 \chi^2}{\partial a_i \partial a_j} > 0$ (improbable)

Saddle point: some $\frac{\partial^2 \chi^2}{\partial a_i \partial a_j} > 0$ (probable)

An average global minimum: in properly chosen coordinates, $\frac{\partial^2 \chi^2}{\partial z_i \partial z_j} > 0$ for dominant coordinate components

Y. Bengio, 2019 Turing lecture (YouTube)

Many dimensions introduce major difficulties with finding a global minimum…

The Loss Surfaces of Multilayer Networks
A. Choromanska, M. Henaff, M. Mathieu, G. Ben Arous, Y. LeCun PMLR 38:192-204, 2015

An important question concerns the distribution of critical points (maxima, minima, and saddle points) of such functions. Results from random matrix theory applied to spherical spin glasses have shown that these functions have a combinatorially large number of saddle points. Loss surfaces for large neural nets have many local minima that are essentially equivalent from the point of view of the test error, and these minima tend to be highly degenerate, with many eigenvalues of the Hessian near zero.

We empirically verify several hypotheses regarding learning with large-size networks:

- For large-size networks, most local minima are equivalent and yield similar performance on a test set.
- The probability of finding a "bad" (high value) local minimum is non-zero for small-size networks and decreases quickly with network size.
- Struggling to find the global minimum on the training set (as opposed to one of the many good local ones) is not useful in practice and may lead to overfitting.

Many dimensions introduce major difficulties with finding a global minimum…

…as well as with representative exploration of uncertainties

The Big Data Paradox in vaccine uptake

Article

Unrepresentative big surveys significantly overestimated US vaccine uptake

https://doi.org/10.1038/s41586-021-04198-4

Received: 18 June 2021

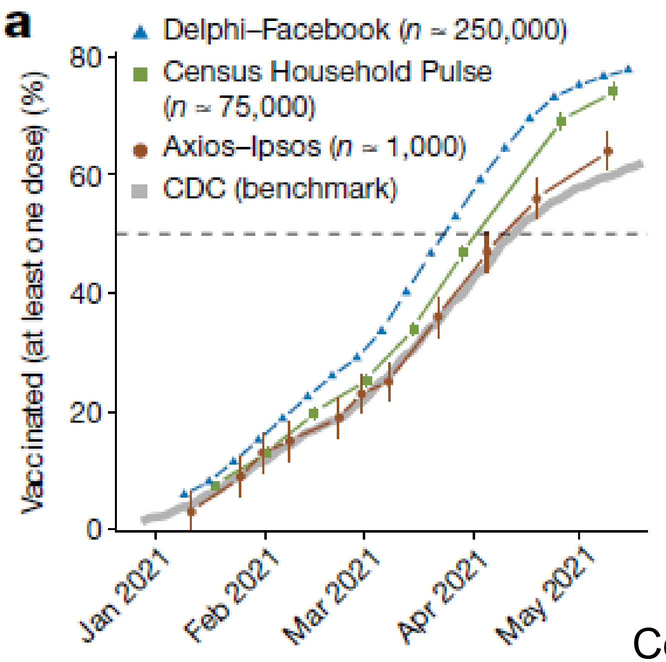
Accepted: 29 October 2021

Published online: 8 December 2021

Check for updates

Valerie C. Bradley[1,6], Shiro Kuriwaki[2,6], Michael Isakov[3], Dino Sejdinovic[1], Xiao-Li Meng[4] & Seth Flaxman[5,6]

Surveys are a crucial tool for understanding public opinion and behaviour, and their accuracy depends on maintaining statistical representativeness of their target populations by minimizing biases from all sources. Increasing data size shrinks confidence intervals but magnifies the effect of survey bias: an instance of the Big Data Paradox[1]. Here we demonstrate this paradox in estimates of first-dose COVID-19 vaccine uptake in US adults from 9 January to 19 May 2021 from two large surveys: Delphi–Facebook[2,3] (about 250,000 responses per week) and Census Household Pulse[4] (about 75,000 every two weeks). In May 2021, Delphi–Facebook overestimated uptake by 17 percentage points (14–20 percentage points with 5% benchmark imprecision) and Census Household Pulse by 14 (11–17 percentage points with 5% benchmark imprecision), compared to a retroactively updated benchmark the Centers for Disease Control and Prevention published on 26 May 2021. Moreover, their large sample sizes led to miniscule margins of error on the incorrect estimates. By contrast, an Axios–Ipsos online panel[5] with about 1,000 responses per week following survey research best practices[6] provided reliable estimates and uncertainty quantification. We decompose observed error using a recent analytic framework[1] to explain the inaccuracy in the three surveys. We then analyse the implications for vaccine hesitancy and willingness. We show how a survey of 250,000 respondents can produce an estimate of the population mean that is no more accurate than an estimate from a simple random sample of size 10. Our central message is that data quality matters more than data quantity, and that compensating the former with the latter is a mathematically provable losing proposition.



Nature v. 600 (2021) 695

Courtoy et al., PRD 107 (2023) 034008

Possible contents for the PDF4LHC document

Details in backup slides

1. Importance of RRR = Reproducibility, Replicability, Rigor
2. Universal factors affecting replicability
3. Recommended practices for improving replicability and reproducibility in the LHC analyses
4. A PDF4LHC wish list for systematic uncertainties
5. “The PDF4LHC challenge”: a scheme to reward PDF fitting groups for quoting PDF uncertainties that are both precise and replicable

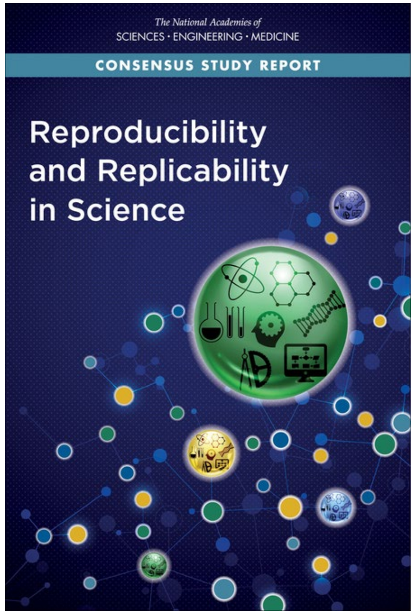
Backup

Reproducibility, **R**eplicability, **R**igor: definitions

Reproducibility is obtaining consistent results using the same input data; computational steps, methods, and code; and conditions of analysis.

Replicability is obtaining consistent results across studies aimed at answering the same scientific question, each of which has obtained its own data.

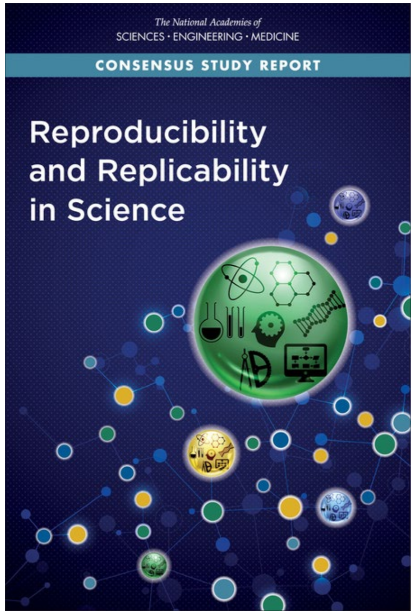
Rigor -- the strict application of the scientific method to ensure robust and unbiased experimental design -- makes replication of a study more likely



Definitions adopted from "*REPRODUCIBILITY AND REPLICABILITY IN SCIENCE",* Conclusion 3.1
National Academy of Sciences, 2019, https://doi.org/10.17226/25303

Universal factors affecting replicability

- complexity of the system under study;
- understanding of the number and relations among variables within the system under study;
- ability to control the variables;
- levels of noise within the system (or signal to noise ratios);
- mismatch of scale of the phenomena and the scale at which it can
- be measured;
- stability across time and space of the underlying principles;
- fidelity of the available measures to the underlying system under study (e.g., direct or indirect measurements);
- prior probability (pre-experimental plausibility) of the scientific hypothesis.



From *"REPRODUCIBILITY AND REPLICABILITY IN SCIENCE"*
National Academy of Sciences, 2019, https://doi.org/10.17226/25303

Recommendations for improving replicability of studies

All researchers should include a clear, specific, and complete description of how a reported result was reached, … including

- a clear description of all methods, instruments, materials, procedures, measurements, and other variables involved in the study;
- a clear description of the analysis of data and decisions for inclusion/exclusion of some data;
- for results that depend on statistical inference, a description of the analytic decisions and when these decisions were made and whether the study is exploratory or confirmatory;
- a discussion of the expected constraints on generality, such as which methodological features the authors think could be varied without affecting the result and which must remain constant;
- reporting of precision or statistical power; and
- a discussion of the uncertainty of the measurements, results, and inferences.

Researchers who use statistical inference analyses should be trained to use them properly.

Funding agencies and organizations should consider investing in R & D of open-source, usable tools and infrastructure that support reproducibility for a broad range of studies across different domains in a seamless fashion**.**

Journals should consider ways to ensure computational reproducibility for studies to the extent it is ethically and legally possible.

From *“REPRODUCIBILITY AND REPLICABILITY IN SCIENCE” ,* https://doi.org/10.17226/25303

PDF wish list for systematic uncertainties

A proposal

Fundamental issues in propagating systematic uncertainties. Some possible remedies:

1. More complete representations for experimental likelihoods that do not need reverse engineering
2. Agreed-upon nomenclature for leading syst. sources
3. Is reducing dimensionality of published correlation matrices advisable? Is there a standard for it? E.g., fewer nuisance parameters; collect less relevant/certain nuisance parameters into one uncorrelated error; etc.
4. Mathematical consistency of covariance/correlation matrices (see Z. Kassabov et al.)
5. How do different implementations of syst. errors affect pulls on PDFs? L_2 sensitivities to nuisance parameters
6. …

aleatory vs. **epistemic** uncertainties

Statistical uncertainty propagated from experiments — reduced by increasing data size

Uncertainty due to lack of knowledge or incomplete models —bias (may be reduced by analysis improvements)

Epistemic PDF uncertainty is important in W boson mass and α_s measurements

ATLAS-CONF-2023-004

PDF-Set	p_T^ℓ [MeV]	m_T [MeV]	combined [MeV]
CT10	$80355.6^{+15.8}_{-15.7}$	$80378.1^{+24.4}_{-24.8}$	$80355.8^{+15.7}_{-15.7}$
CT14	$80358.0^{+16.3}_{-16.3}$	$80388.8^{+25.2}_{-25.5}$	$80358.4^{+16.3}_{-16.3}$
CT18	$80360.1^{+16.3}_{-16.3}$	$80382.2^{+25.3}_{-25.3}$	$80360.4^{+16.3}_{-16.3}$
MMHT2014	$80360.3^{+15.9}_{-15.9}$	$80386.2^{+23.9}_{-24.4}$	$80361.0^{+15.9}_{-15.9}$
MSHT20	$80358.9^{+13.0}_{-16.3}$	$80379.4^{+24.6}_{-25.1}$	$80356.3^{+14.6}_{-14.6}$
NNPDF3.1	$80344.7^{+15.6}_{-15.5}$	$80354.3^{+23.6}_{-23.7}$	$80345.0^{+15.5}_{-15.5}$
NNPDF4.0	$80342.2^{+15.3}_{-15.3}$	$80354.3^{+22.3}_{-22.4}$	$80342.9^{+15.3}_{-15.3}$

Table 2: Overview of fitted values of the W boson mass for different PDF sets. The reported uncertainties are the total uncertainties.

ATLAS-CONF-2023-015

The statistical analysis for the determination of $\alpha_s(m_Z)$ is performed with the xFitter framework [60]. The value of $\alpha_s(m_Z)$ is determined by minimising a χ^2 function which includes both the experimental uncertainties and the theoretical uncertainties arising from PDF variations:

$$\chi^2(\beta_{\rm exp}, \beta_{\rm th}) = \sum_{i=1}^{N_{\rm data}} \frac{\left(\sigma_i^{\rm exp} + \sum_j \Gamma_{ij}^{\rm exp}\beta_{j,\rm exp} - \sigma_i^{\rm th} - \sum_k \Gamma_{ik}^{\rm th}\beta_{k,\rm th}\right)^2}{\Delta_i^2} + \sum_j \beta_{j,\rm exp}^2 + \sum_k \beta_{k,\rm th}^2 \,. \quad (1)$$

profiling of CT and MSHT PDFs requires to include a tolerance factor $T^2 > 10$ as in the ePump code

[T.J. Hou et al., 1912.10053, Appendix F]

Also the next slide.

Augmented likelihood for PDFs with global tolerance

1. Start by defining the correspondence between $\Delta\chi^2$ and cumulative probability level: 68% c.l. ⇔ $\Delta\chi^2 = T^2$.
2. Write the **augmented** likelihood density for this definition:

$$P(D_i|T_i) \propto e^{-\chi^2/(2T^2)}$$

3. When profiling 1 new experiment with the prior imposed on PDF nuisance parameters $\lambda_{\alpha,th}$:

$$\chi^2(\vec{\lambda}_{\rm exp}, \vec{\lambda}_{\rm th}) = \sum_{i=1}^{N_{pt}} \frac{\left[D_i + \sum_\alpha \beta^{\rm exp}_{i,\alpha}\lambda_{\alpha,\rm exp} - T_i - \sum_\alpha \beta^{\rm th}_{i,\alpha}\lambda_{\alpha,\rm th}\right]^2}{s_i^2} + \sum_\alpha \lambda^2_{\alpha,\rm exp} + \sum_\alpha T^2 \lambda^2_{\alpha,\rm th}. \qquad \beta^{\rm th}_{i,\alpha} = \frac{T_i(f^+_\alpha) - T_i(f^-_\alpha)}{2},$$

new experiment

priors on expt. systematics and PDF params

4. Alternatively, we can reparametrize $\chi^{2'} \equiv \chi^2/T^2$, so that 68% c.l. ⇔ $\Delta\chi^{2'} = 1$. We have

$$P(D_i|T_i) \propto e^{-\chi^{2'}/2}$$

$$\chi^{2'}(\vec{\lambda}_{\rm exp}, \vec{\lambda}_{\rm th}) = \sum_{i=1}^{N_{pt}} \frac{\left[D_i + \sum_\alpha \beta^{\rm exp}_{i,\alpha}\lambda_{\alpha,\rm exp} - T_i - \sum_\alpha \beta^{\rm th}_{i,\alpha}\lambda_{\alpha,\rm th}\right]^2}{s_i^2\; T^2} + \sum_\alpha \frac{\lambda^2_{\alpha,\rm exp}}{T^2} + \sum_\alpha \lambda^2_{\alpha,\rm th}.$$

consistent redefinition

5. Inconsistent redefinitions:

$$\chi^{2'}(\vec{\lambda}_{\rm exp}, \vec{\lambda}_{\rm th}) = \sum_{i=1}^{N_{pt}} \frac{\left[D_i + \sum_\alpha \beta^{\rm exp}_{i,\alpha}\lambda_{\alpha,\rm exp} - T_i - \sum_\alpha \beta^{\rm th}_{i,\alpha}\lambda_{\alpha,\rm th}\right]^2}{s_i^2} + \sum_\alpha \lambda^2_{\alpha,\rm exp} + \sum_\alpha \lambda^2_{\alpha,\rm th}.$$

and $P(D_i|T_i) \propto e^{-\chi^{2'}/2}$

or $P(D_i|T_i) \propto e^{-\chi^{2'}/(2T^2)}$

[equivalent to $s_i \to s_i/T$ or $\lambda_{\alpha,th} \to \lambda_{\alpha,th}T$ without $\beta_{i,\alpha,th} \to \beta_{i,\alpha,th}/T$]

Why **augmented** likelihood?

The term is accepted in lattice QCD to indicate that the log-likelihood contains **prior terms**

$$\chi^2(\vec{\lambda}_{\rm exp}, \vec{\lambda}_{\rm th}) = \sum_{i=1}^{N_{pt}} \frac{\left[D_i + \sum_\alpha \beta_{i,\alpha}^{\rm exp} \lambda_{\alpha,\rm exp} - T_i - \sum_\alpha \beta_{i,\alpha}^{\rm th} \lambda_{\alpha,\rm th}\right]^2}{s_i^2} + \sum_\alpha \lambda_{\alpha,\rm exp}^2 + \sum_\alpha T^2 \lambda_{\alpha,\rm th}^2.$$

new experiment

priors on expt. systematics
and PDF params

After minimization w.r.t. to $\lambda_{\alpha,exp}$, $\lambda_{\alpha,th}$, the prior terms are **hidden** inside the covariance matrix:

$$\chi^2 = \sum_{i,j}^{N_{pt}} (T_i - D_i)(\text{cov}^{-1})_{ij}\left(T_j - D_j\right)$$

The usual χ^2 definition therefore contains a **prior** component, which may be handled differently by the various groups

Complexity and PDF tolerance

- **Bad news:** The tolerance puzzle is *intractable* in very complex fits
  - In a fit with N_{par} free parameters, the minimal number of PDF replicas to estimate the expectation values for ∀ χ^2 function grows as $N_{min} \geq 2^{N_{par}}$
  - Example: $N_{min} > 10^{30}$ for $N_{par} = 100$

[Sloan, Wo´zniakowski, 1997]

[Hickernell, MCQMC 2016, 1702.01487]

Good news: expectation values **for typical QCD observables** can be estimated with fewer replicas by reducing dimensionality of the problem or a targeted sampling technique.

Example: a "**hopscotch scan**", see 2205.10444