Prior Specification in Bayesian Estimation Affects a Model's Fitting Propensity

Sonja D. Winter, PhD

Better Decisions Through Comprehensive Statistical Model Evaluation

April 14, 2023: NCME 2023, Chicago

Missouri Prevention Science Institute
University of Missouri

Model Fitting, according to frequentists

new evidence
(likelihood)

$$\frac{p(data\,|\theta)}{p(data)}$$

number of
parameters

configuration
of the model

Model Fitting, according to Bayesians

existing information
(prior)

new evidence
(likelihood)

updated information
(posterior)

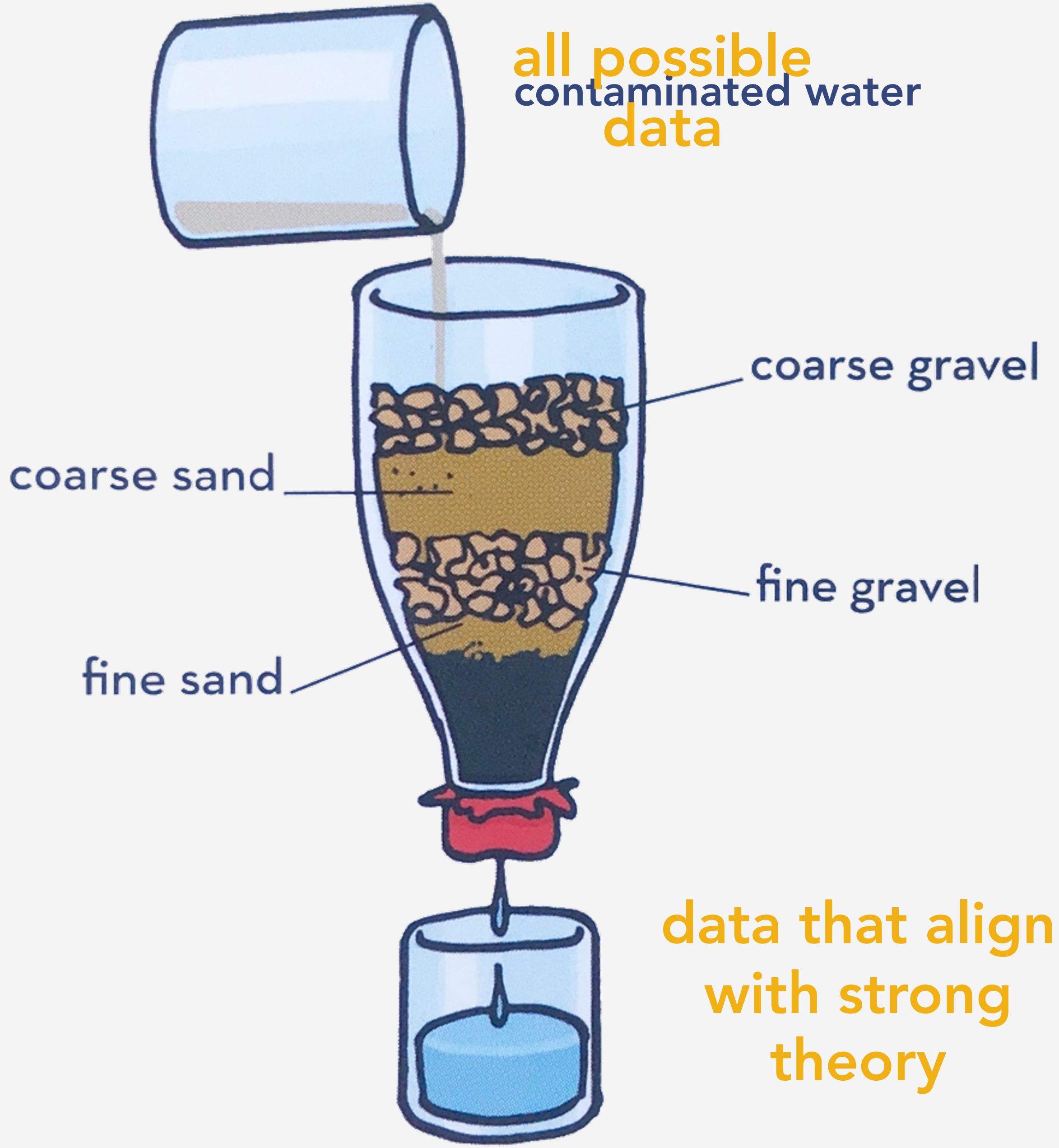
$$p(\theta) \times \frac{p(data\,|\theta)}{p(data)} = p(\theta|data)$$

Can Priors make a Model More Selective?

- When a model ‘fits well’, it does not tell us anything about the a priori likelihood of that model fitting *any plausible data*
  - Some models have a worryingly high tendency to fit any data patterns (i.e., high fitting propensity; FP; Preacher, 2006)
  - Without constraining such models, finding good fit is ‘nearly meaningless’ (Roberts & Pashler, 2000)

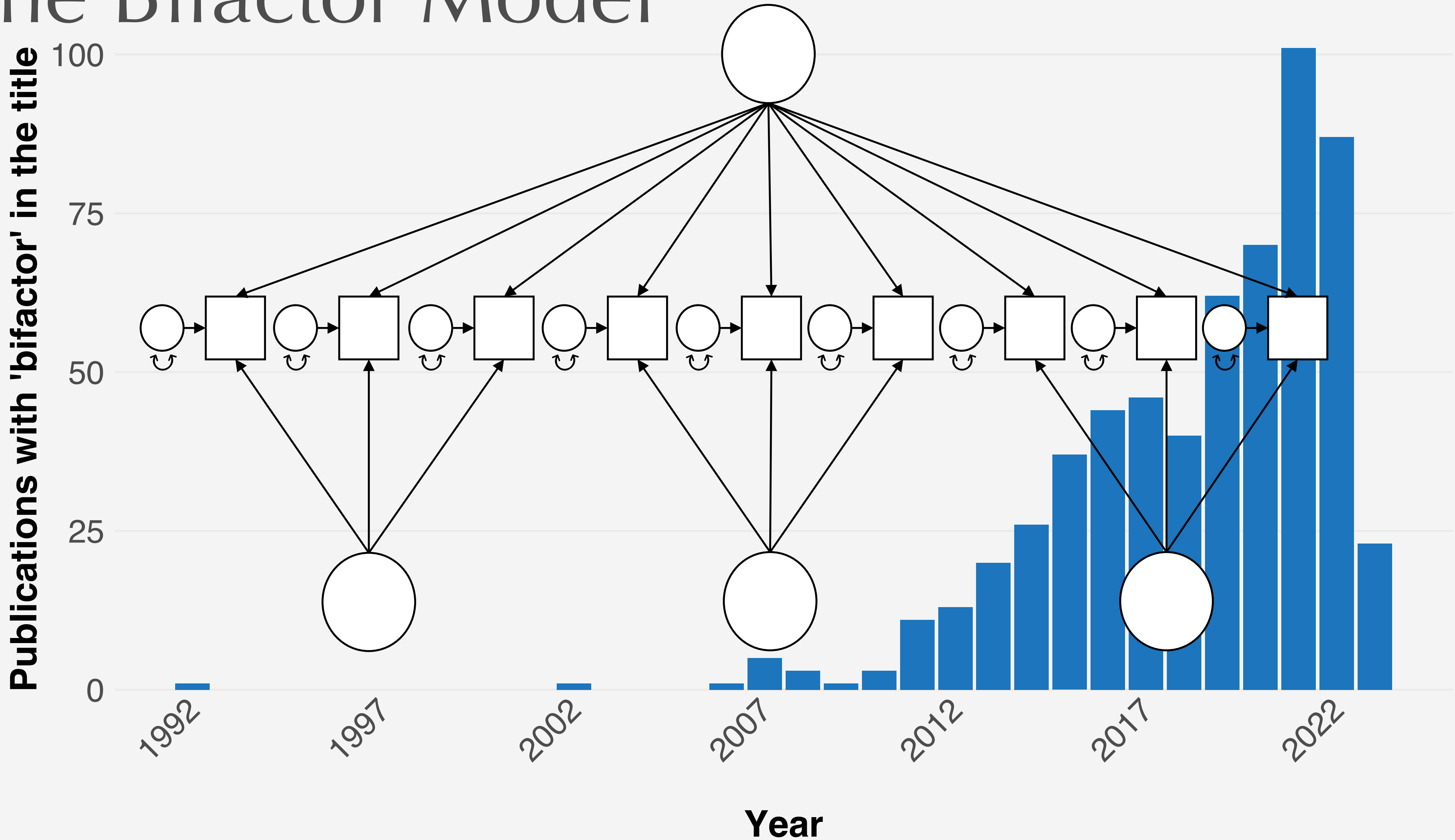
- Can we return meaning to good model fit?
- Can we use a series of increasingly fine-grained prior specifications to ensure that our model fits well only to data that align to our theory?
  - Extending work by Vanpaemel (2009; Vanpaemel & Lee, 2012)

Priors as Filters



Source: Bear Grylls Survival Academy

Example of a Model with high FP: The Bifactor Model



Criticism of the well-fitting Bifactor Model

- “Indiscriminate use of the bifactor model without proper regard for theory is highly questionable.” (Thomas, 2012, p. 108)
- “[W]e caution against the adoption of a theoretical model that is built on a methodological house of cards.” (Watts et al., 2020, p. 318)
- “[T]he bifactor model has an undesirable tendency to fit any possible data” (Bonifay & Cai, 2017, p. 481)
- “[T]he mistaken inference of bifactor superiority seems to be driven by the general dimension’s erroneous accommodation of misspecifications through capturing theoretically unexplained variance and repackaging it as common variance, even though it is not.” (Greene et al., p. 756)

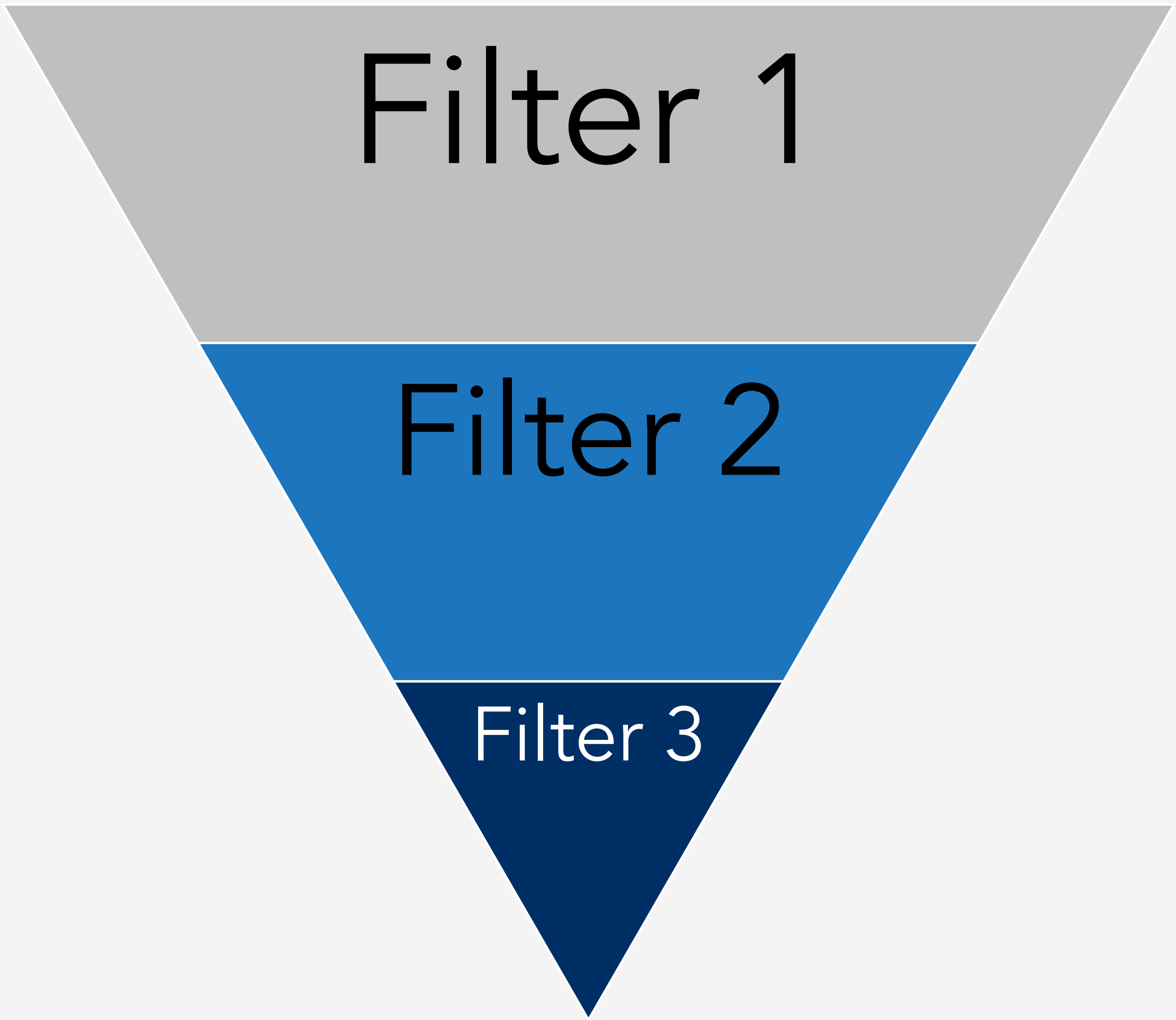
Theory-Informed Constraints of the Bifactor Model

Watts and colleagues (2019) proposed two theory-informed constraints for psychopathology bifactor model parameter values:

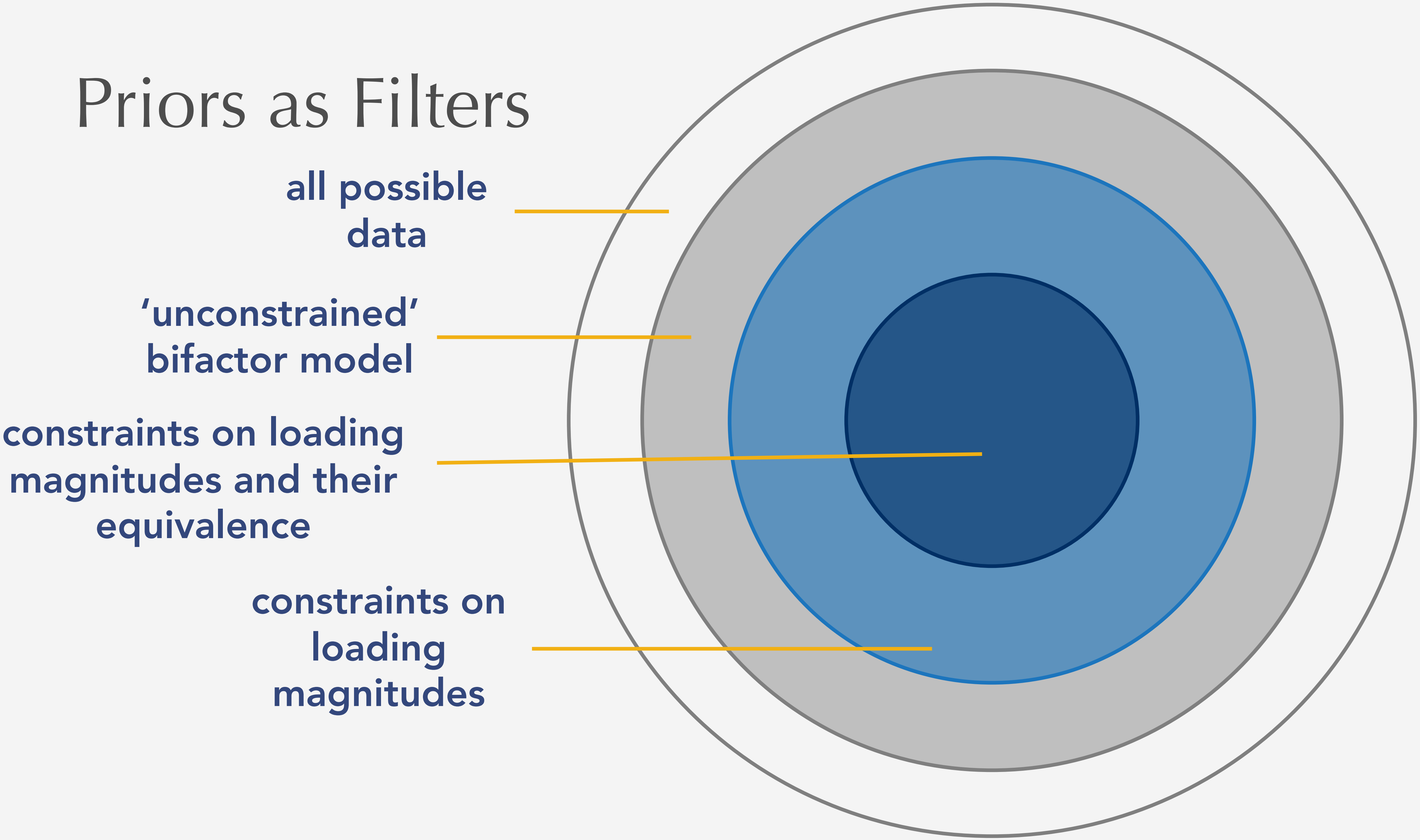
1. A bifactor model should produce reliable specific factors that are well represented by their constituent indicators.
2. If the general factor in a bifactor model reflects broad liability for psychopathology, it should be relatively equally represented by its constituent indicators.

We can translate these constraints to prior *filters*

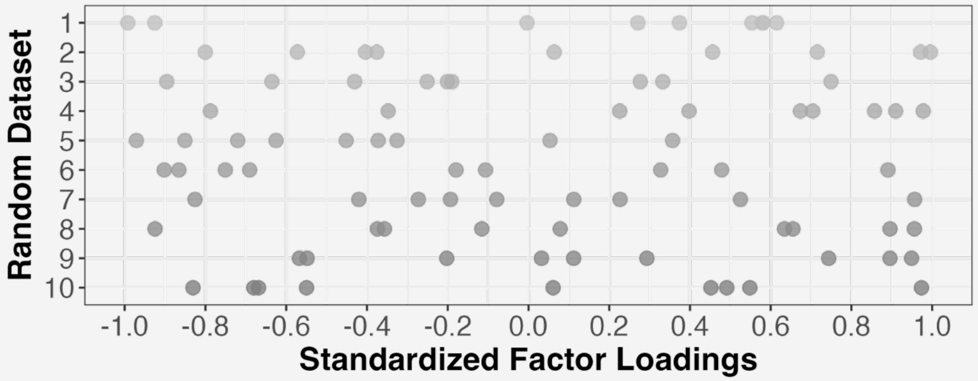
Priors as Filters

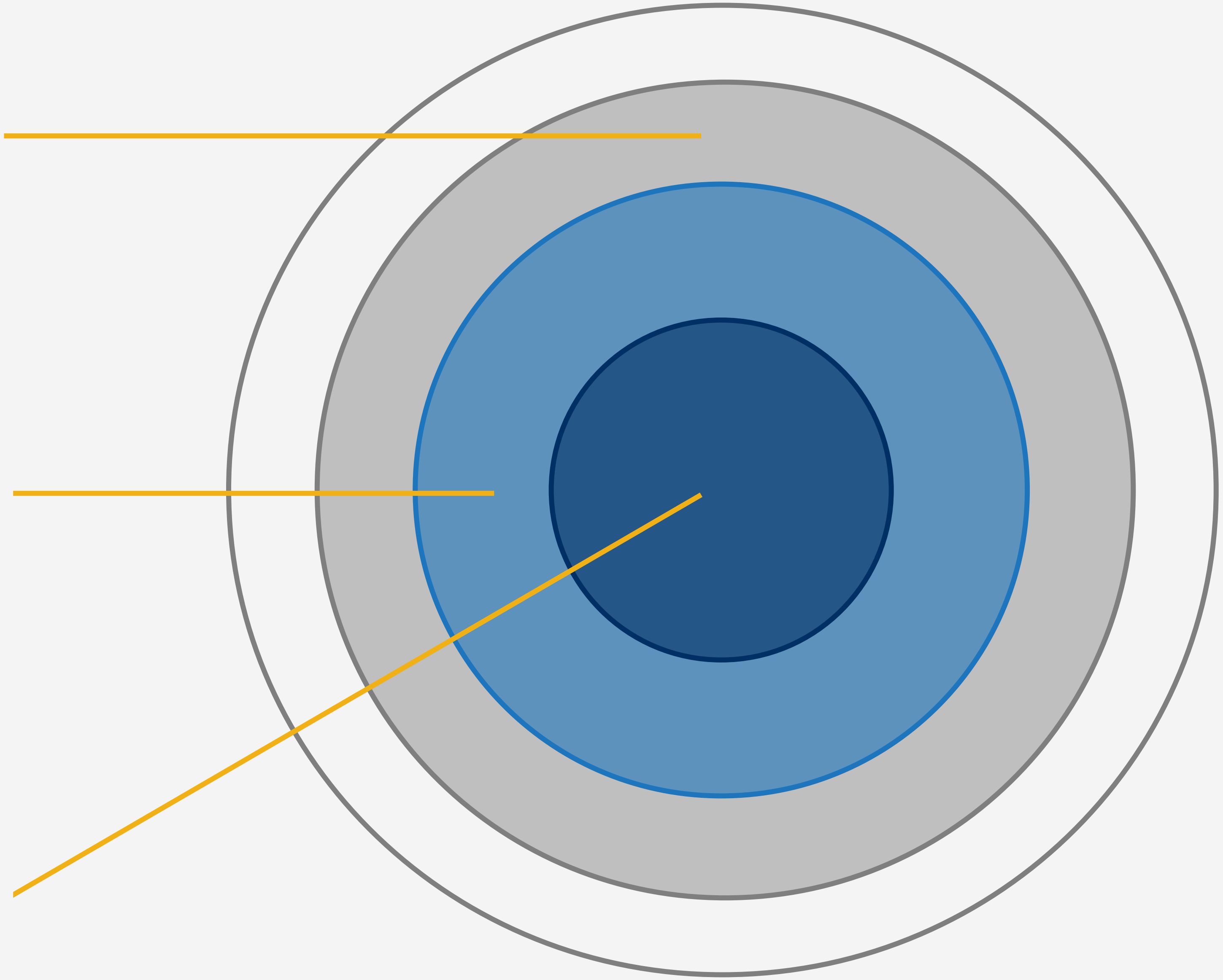


Priors as Filters



Filter 1: $\lambda \sim \mathrm{N}(\mu = 0, \sigma = 10)$



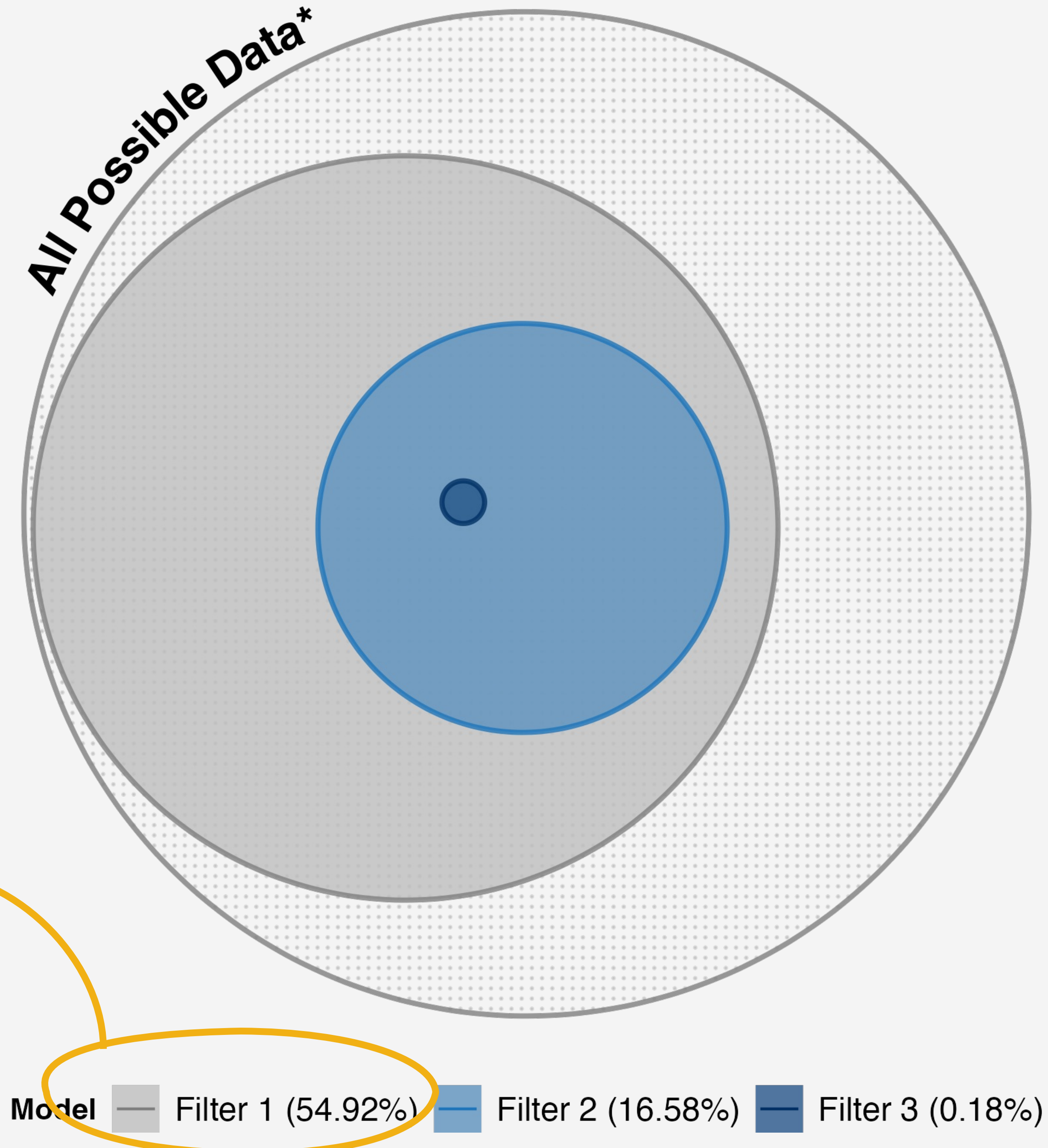


FP Analysis to Test Priors as Filters

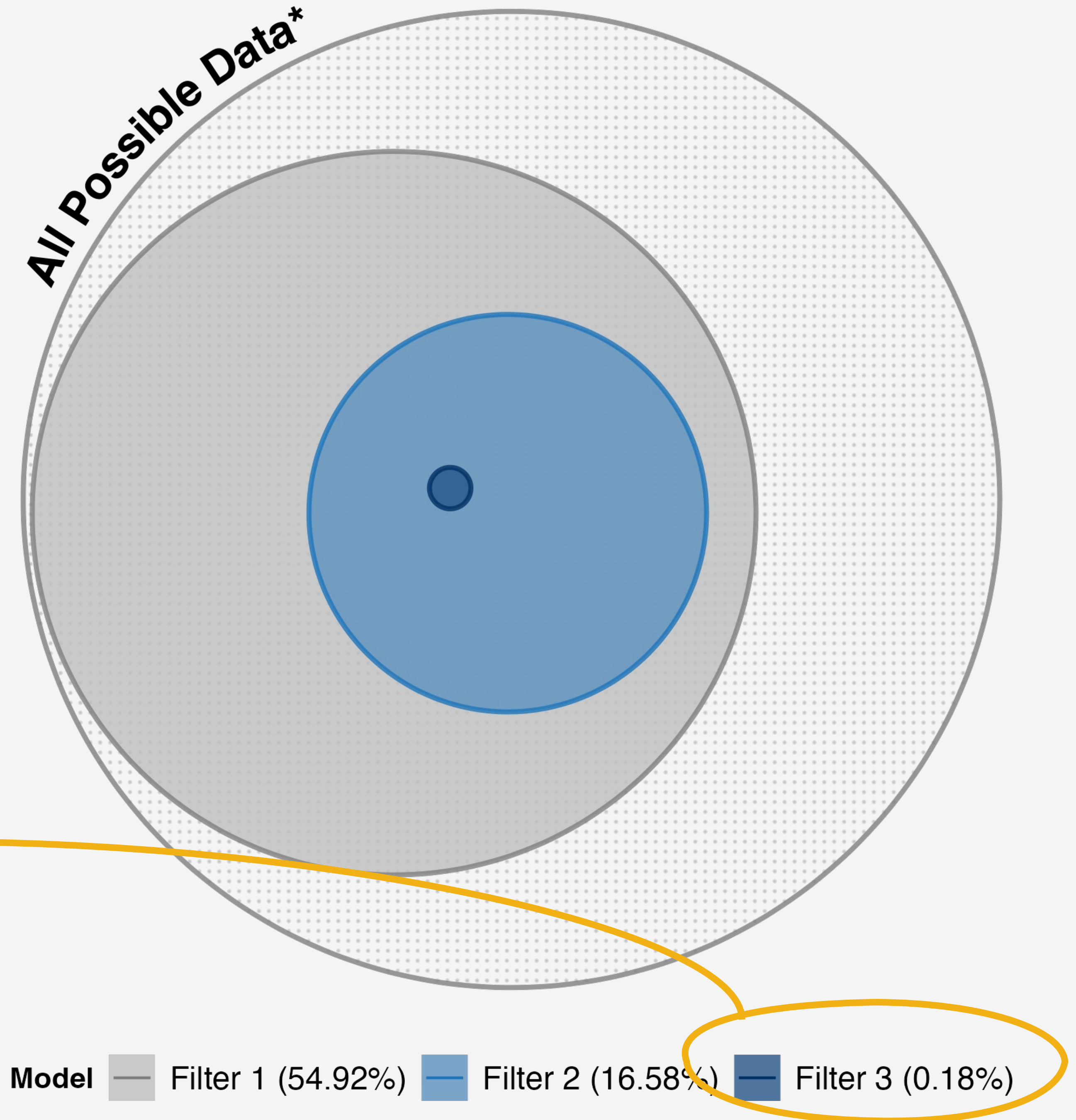
1. Generate 5000 random data sets of N = 500 with 15 variables
   - **ockhamSEM** R package (Falk & Muthukrishna, 2021)
2. Fit bifactor model with diffuse priors (Filter 1)
   - **blavaan** R package (Merkle et al., 2021)
3. Assess model fit
   - Bayesian SRMR ≤ .12
     - This index does not have same cutoff guidelines as frequentist SRMR
4. Good fit? → Apply Filter 2 Priors
5. Still good fit with Filter 2? → Apply Filter 3 Priors

FP Analysis Results

55% of all possible data got through the 'unconstrained' *bifactor model filter*



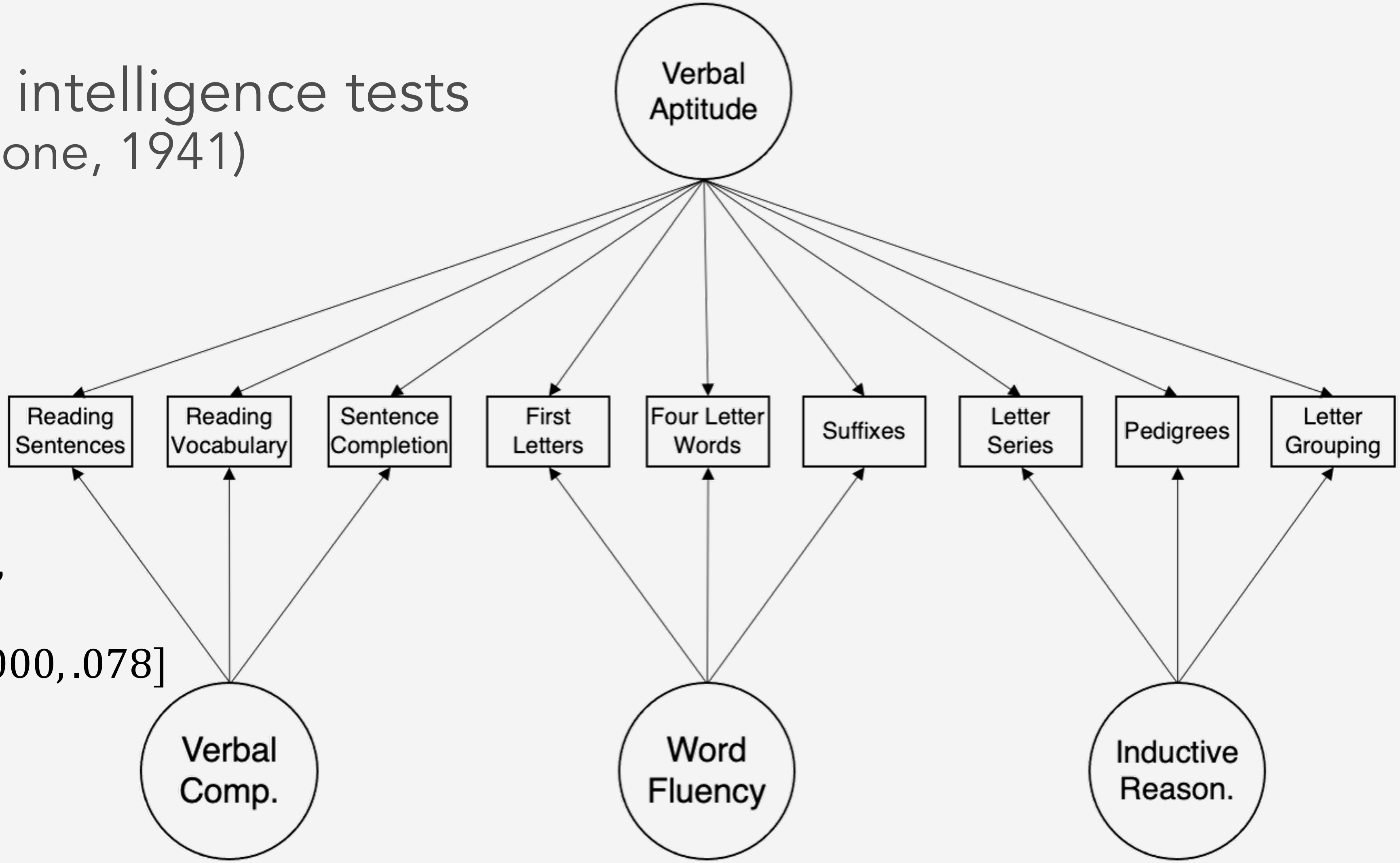
FP Analysis Results



Only 0.18% of all possible data got through the *most fine-grained filter (well-represented + equally well-represented)*

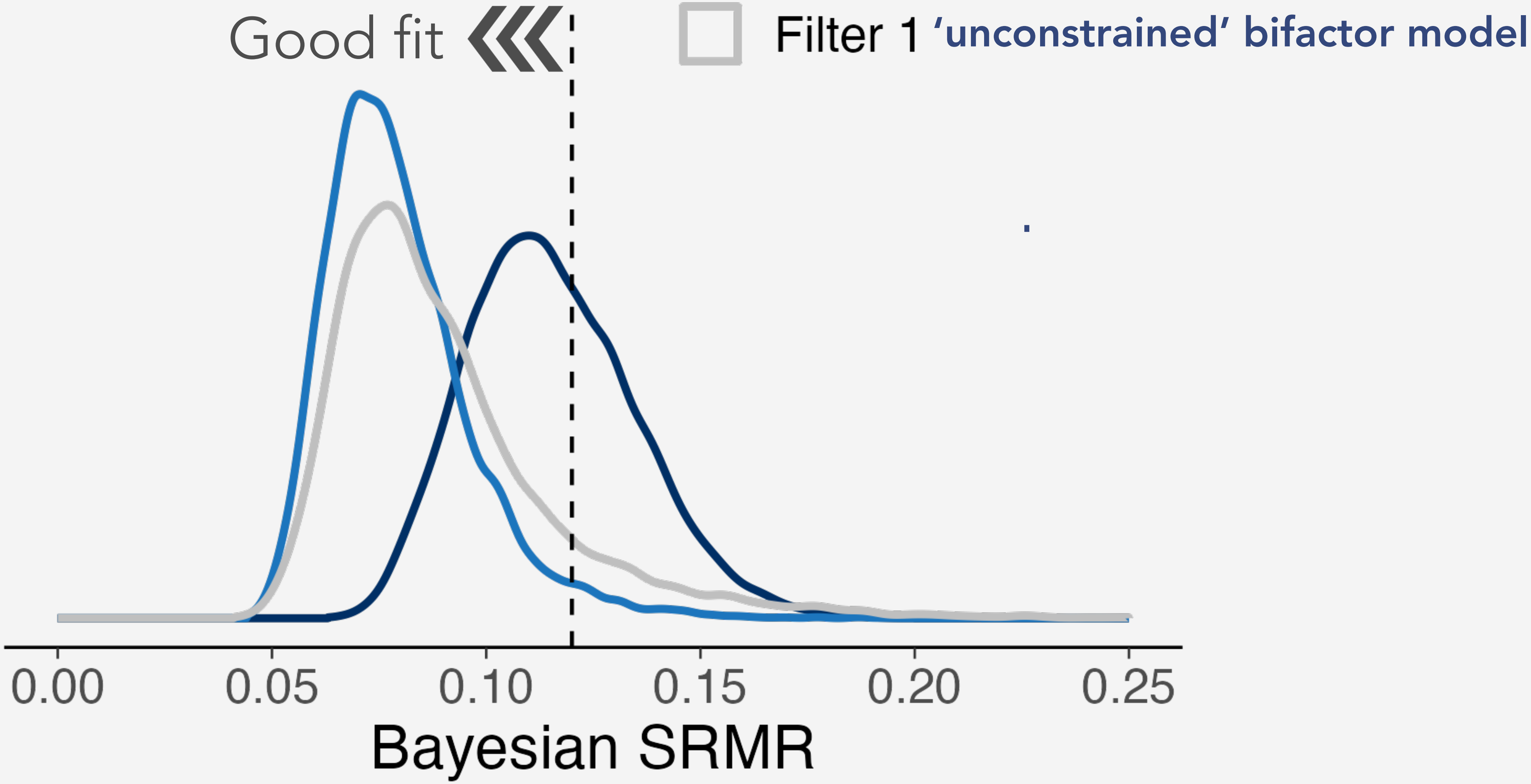
Applying Filters to a Single Data Set

- Thurstone verbal intelligence tests (Thurstone & Thurstone, 1941)
- N = 213



Frequentist Model Fit:
$\chi^2\,(18) = 24.33, p = .144,$
$CFI = .994,$
$RMSEA = .041, 90\%\ CI\ [.000, .078]$

Applying Constraints to Example



Conclusions & Implications

We can use Bayesian priors to constrain statistical the bifactor model!

- This general procedure can be applied to any statistical model/type of parameter
- We can use priors to constrain the complexity of any model for which priors can be specified
  - Return meaning to *good model fit*
- We need more insight into:
  - How to specify proper priors for specific sample and model sizes
  - What goodness of fit indices are best

Thank you!

sdwinter@missouri.edu

@winterstat

Questions? Feedback?

But is Good Fit Meaningful?

Verbal Aptitude	
Reading Sentences	0.766
Reading Vocabulary	0.789
Sentence Completion	0.752
First Letters	0.607
Four Letter Words	0.596
Suffixes	0.570
Letter Series	0.566
Pedigrees	0.661
Letter Grouping	0.529

Verbal Comprehension	
Reading Sentences	0.487
Reading Vocabulary	0.451
Sentence Completion	0.404

Word Fluency	
First Letters	0.613
Four Letter Words	0.505
Suffixes	0.393

Inductive Reasoning	
Letter Series	0.726
Pedigrees	0.246
Letter Grouping	0.408

References

- Bonifay, W., & Cai, L. (2017). On the Complexity of Item Response Theory Models. *Multivariate Behavioral Research*, *52*(4), 465–484.
- Falk, C. F., & Muthukrishna, M. (2021). *ockhamSEM: Tools for studying fit propensity in structural equation modeling*. R package version 0.1.2.
- Greene, A. L., Eaton, N. R., Li, K., Forbes, M. K., Krueger, R. F., Markon, K. E., ... & Kotov, R. (2019). Are fit indices used to test psychopathology structure biased? A simulation study. *Journal of abnormal psychology*, *128*(7), 740.
- Levy, R. (2011). Bayesian Data-Model Fit Assessment for Structural Equation Modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *18*(4), 663–685.
- Lewandowski, D., Kurowicka, D., & Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis, 100*(9), 1989–2001
- Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology*, *46*, 806–834.
- Merkle E.C., Fitzsimmons E., Uanhoro J., Goodrich B. (2021). Efficient Bayesian Structural Equation Modeling in Stan. *Journal of Statistical Software, 100*(6), 1-22.
- Murray, A. L., & Johnson, W. (2013). The limitations of model fit in comparing the bi-factor versus higher-order models of human cognitive ability structure. *Intelligence*, *41*(5), 407-422.
- Myung, I. J., Pitt, M. A., & Kim, W. (2005). Model evaluation, testing and selection. In K. Lambert & R. Goldstone (Eds.), *Handbook of Cognition*. Sage Publications Ltd.
- Preacher, K. J. (2006). Quantifying Parsimony in Structural Equation Modeling. *Multivariate Behavioral Research*, *41*(3), 227–259.
- Stan Development Team (2022). *RStan: the R interface to Stan*. R package version 2.26.13.
- Thomas, M. L. (2012). Rewards of bridging the divide between measurement and clinical theory: demonstration of a bifactor model for the Brief Symptom Inventory. *Psychological assessment*, *24*(1), 101
- Vanpaemel, W. (2009). Measuring model complexity with the prior predictive. *Advances in neural information processing systems, 22.*
- Vanpaemel, W. (2020). Strong theory testing using the prior predictive and the data prior. *Psychological Review, 127*(1), 136–145.
- Vanpaemel, W., & Lee, M. D. (2012). Using priors to formalize theory: Optimal attention and the generalized context model. *Psychonomic Bulletin & Review*, *19*(6), 1047-1056.
- Watts, A. L., Lane, S. P., Bonifay, W., Steinley, D., & Meyer, F. A. (2020). Building theories on top of, and not independent of, statistical models: The case of the p-factor. *Psychological inquiry*, *31*(4), 310-320.
- Watts, A. L., Poore, H. E., & Waldman, I. D. (2019). Riskier tests of the validity of the bifactor model of psychopathology. *Clinical Psychological Science*, *7*(6), 1285-1303.