

REPLY TO PAUL E. MEEHL:
MACHINE LEARNING STRUCTURAL
EQUATION MODELING AND
FALSIFICATORY DATA ANALYSIS

Michael Shiuming Truong

York University

September 29, 2025



Artworks generated using Google (2025a, 2025b). Contact: emstruonger@gmail.com

TODAY'S PLAN

- PART 1: Manuscript 1: FDA
- PART 2: Manuscript 2: ML-SEM
- PART 3: Manuscript 3: ML-SEM and FDA
- PART 4: General Conclusion & Big Picture
- PART 5: Appendix

What is Today's Destination?

REPLY TO PAUL E. MEEHL

Risky tests of specific predictions made by theories is an important idea, but is beset by too many challenging problems.

Assertion: the zone of possibility is far larger than the zone of impossibility.

Instead, let's do risky tests of specific 'anti-predictions' made by theories.

WHAT ARE THE GENERAL EXPECTED OUTCOMES?



WHAT COULD BE THE CENTRAL DISEASE?



Assertion: You cannot replace a very poor theory with 'no theory'.

HOW MIGHT THIS CENTRAL DISEASE BE CURED?





+

Manuscript 1: An Elaboration: Falsificatory Data Analysis



TABLE OF CONTENTS

MODERN PROBLEMS OF PSYCHOLOGICAL SCIENCE

POTENTIAL SOLUTION: ELABORATING TUKEY'S EDA/CDA DISTINCTION

RESEARCH QUESTION

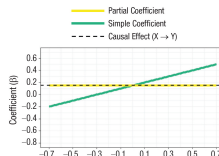
METHODOLOGY

EXPECTED OUTCOME AND BIG PICTURE

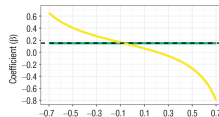
WHAT ARE THE 3 PROBLEMS?

1. Covariate selection is highly impactful (Cinelli et al., 2024; McElreath, 2020; Pearl & Mackenzie, 2018; Wysocki et al., 2022)
2. Multiverse and multi-analyst analyses (Masur & Scharkow, 2020; Simonsohn et al., 2020)
3. The 'Piranha' problem (Tosh et al., 2025)

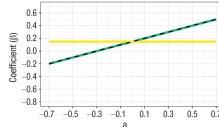
Confounder



Collider



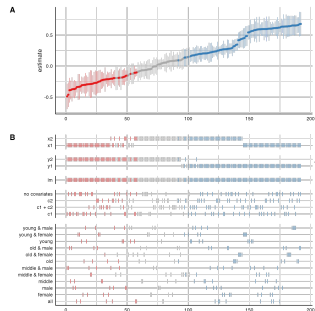
Mediator



PROBLEM 2

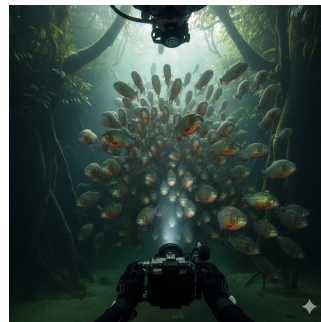
WHAT ARE THE 3 PROBLEMS?

1. Covariate selection is highly impactful (Cinelli et al., 2024; McElreath, 2020; Pearl & Mackenzie, 2018; Wysocki et al., 2022)
2. Multiverse and multi-analyst analyses (Masur & Scharkow, 2020; Simonsohn et al., 2020)
3. The 'Piranha' problem (Tosh et al., 2025)



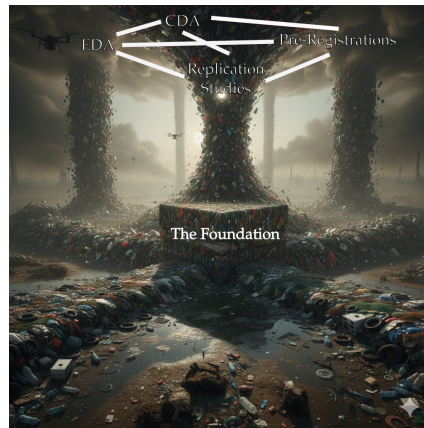
WHAT ARE THE 3 PROBLEMS?

1. **Covariate selection is highly impactful** (Cinelli et al., 2024; McElreath, 2020; Pearl & Mackenzie, 2018; Wysocki et al., 2022)
2. **Multiverse and multi-analyst analyses** (Masur & Scharkow, 2020; Simonsohn et al., 2020)
3. **The 'Piranha' problem** (Tosh et al., 2025)



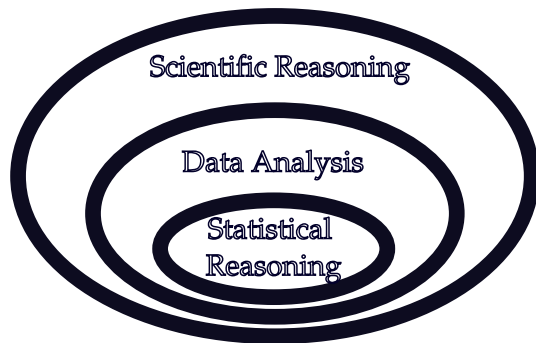
DONC ALORS?

WHY DO THESE PROBLEMS MATTER?



SOLVE BY ELABORATING ON TUKEY'S DISTINCTION: EDA/CDA/FDA

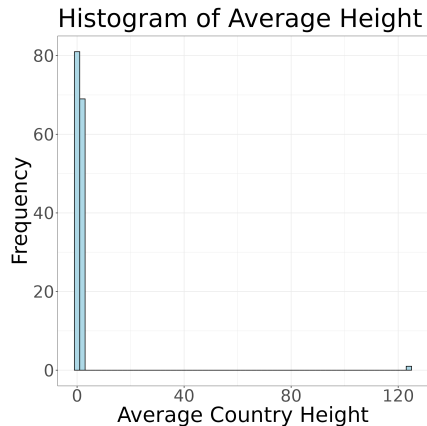
- ▶ EDA $\rightsquigarrow$ What is true?
- ▶ CDA $\rightsquigarrow$ Is what we believe true?
- ▶ FDA $\rightsquigarrow$ What do we refuse to believe is true?
 - ▶ Theory-based rejection of data



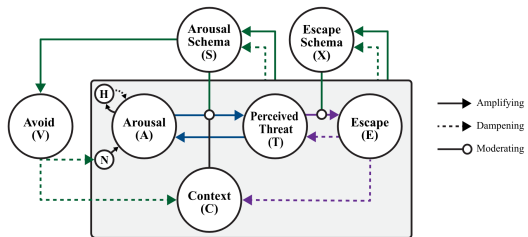
WHAT ARE SOME FDA-SPECIFIC TERMS?

- ▶ Predictors vs Anti-Predictors
 - ▶ Significant effect-of or association-with **predictor** $\rightsquigarrow$ ✓
 - ▶ Significant effect-of or association-with **anti-predictor** $\rightsquigarrow$ ✗
- ▶ Zone of Possibility vs Zone of Impossibility
 - ▶ Theory-based possible and impossible **parameter-space values**
 - ▶ Theory-based possible and impossible **data-space values**

WHAT IS AN EXAMPLE OF AN ANTI-PREDICTOR?



WHAT IS AN EXAMPLE OF THE ZONE OF IMPOSSIBILITY?



This generative models predicts that “individuals who develop panic disorder will ultimately develop maximally severe panic disorder, a **prediction inconsistent with the simple observation that the mean severity of panic disorder typically falls in the moderate range** (Furukawa et al., 2009)”

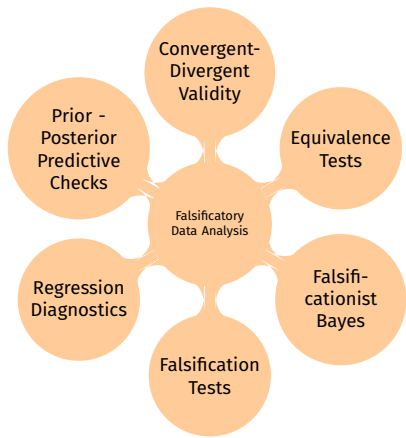
Robinaugh et al. (2024, p.1498)

Failure to account for effects of human learning? (Robinaugh et al., 2024)

WHY SHOULD THIS ELABORATION IMPROVE THINGS?

But will it really improve things? How do we know?

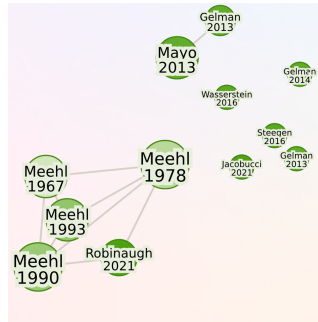
STEP 1: CREATE TERMINOLOGY – EXPLAIN & ORGANIZE CONCEPTS BY RECASTING HISTORICAL EXAMPLES



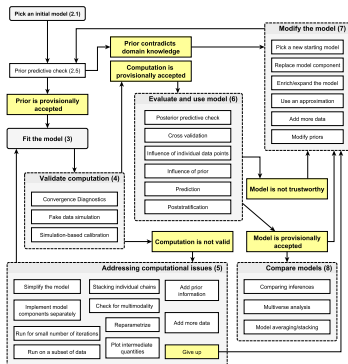
Recasting historical examples

- ▶ Barrett (2017), Nadel and Moscovitch (1997), Robinaugh et al. (2024), and Tversky and Kahneman (1974, 1983)

STEP 2: LOGICAL ANALYSIS GROUNDED IN PHILOSOPHY OF SCIENCE, STATISTICS AND PSYCHOLOGY



STEP 3: CREATE POTENTIAL FLOW CHARTS



(Gelman et al., 2020; Hong et al., 2020)

WHAT IS THE POINT? WHY SHOULD I CARE?

1. Cognitive availability – increase frequency of this kind of scientific reasoning
2. Historically successful precedents
 - ▶ Barrett (2017), Nadel and Moscovitch (1997), Robinaugh et al. (2024), and Tversky and Kahneman (1974, 1983)
3. Logically equivalent to proof by contradiction, *reductio ad absurdum*, etc...
 - ▶ Probabilistic Falsification $\rightsquigarrow$ Deterministic Falsification
 - ▶ But falsification is consensual (McElreath, 2020)



Manuscript 2: A New SEM Trees Variant: IGSCA Trees



TABLE OF CONTENTS

WHAT IS SEM? cSEM? CSA?

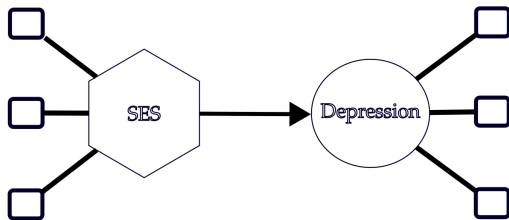
WHAT ARE DECISION TREES? WHAT DO THEY HAVE TO DO WITH SEM/IGSCA TREES?

RESEARCH QUESTION: WHAT IS THE BEST WAY TO DO IGSCA TREES? WHY IS IT NEEDED?

METHODOLOGY

WHAT IS THE POINT? WHY SHOULD I CARE?

WHAT IS SEM? cSEM vs CSA



cSEM

$$\Sigma = \begin{bmatrix} \blacksquare & & \\ & \blacksquare & \\ & & \blacksquare \end{bmatrix}$$

CSA

$$\Sigma = \begin{bmatrix} \blacksquare & \blacksquare & \blacksquare \\ \blacksquare & \blacksquare & \blacksquare \\ \blacksquare & \blacksquare & \blacksquare \end{bmatrix}$$

WHAT IS THE MEASUREMENT MODEL?

See Whiteboard

See Whiteboard

WHAT IS THE WEIGHTED RELATIONS MODEL?

See Whiteboard

See Whiteboard

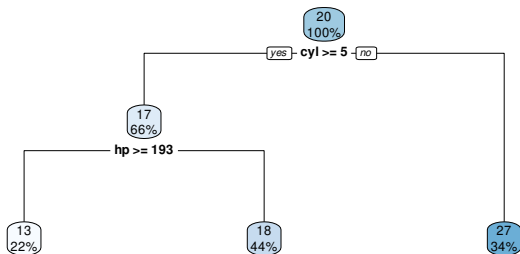


UNDERSTANDING DECISION TREES BY COMPARISON WITH OLS I

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360	175	3.15	3.440	17.02	0	0	3	2
Valiant	18.1	6	225	105	2.76	3.460	20.22	1	0	3	1

UNDERSTANDING DECISION TREES BY COMPARISON WITH OLS II

`rpart::rpart(mpg ~ ., data = mtcars)` $R^2 \approx 0.75$



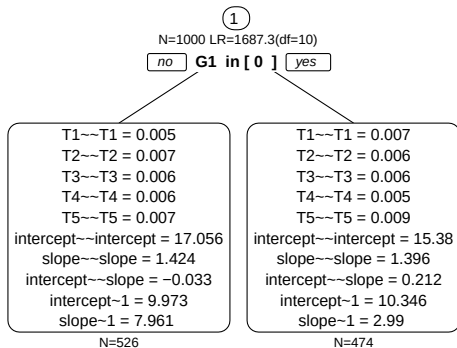
`lm(mpg ~ ., data = mtcars)`
 $R^2 \approx 0.8$

$$\begin{aligned}\widehat{\text{mpg}} = & 12.3 - 0.11(\text{cyl}) + 0.01(\text{disp}) \\ & - 0.02(\text{hp}) + 0.79(\text{drat}) - 3.72(\text{wt}) \\ & + 0.82(\text{qsec}) + 0.32(\text{vs}) + 2.52(\text{am}) \\ & + 0.66(\text{gear}) - 0.2(\text{carb})\end{aligned}$$

HOW ARE DECISION TREES APPLIED TO SEM TREES?

```
semtree::semtree(MODEL, growth.data, predictors = "G1")
```

	T1	T2	T3	T4	T5	G1
1	8.159928	11.85968	15.70002	19.28675	23.04959	0
2	5.513717	11.23166	17.04010	22.51624	28.53268	1
3	9.441117	17.17371	24.76851	32.48777	39.99482	1
4	12.217068	15.01575	17.87970	20.80830	23.51129	0
5	1.400683	10.44168	19.16357	28.01147	36.88147	1
6	9.931001	12.74040	15.57150	18.63229	21.53497	0



HOW CAN I VISUALIZE SEM TREES?

FIGURE 1: Left Daughter Node where $G1 = 1$

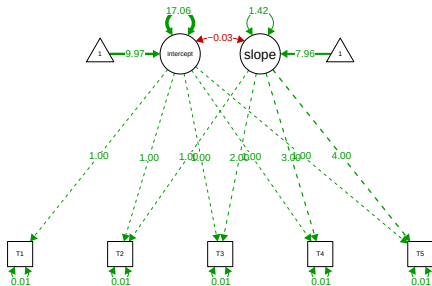
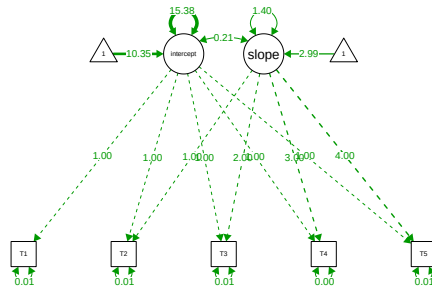


FIGURE 2: Right Daughter Node where $G1 = 0$



WHAT ARE SOME WEAKNESSES OF TRADITIONAL SEM TREES? HOW MIGHT IGSCA TREES ADDRESS THIS GAP?

Traditional SEM Trees inherits the strengths and weaknesses of traditional CSA.

Pros of IGSCA Trees:

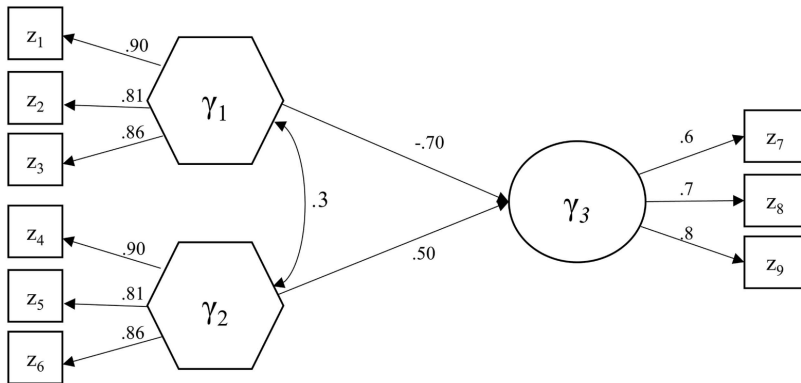
- ▶ No likelihood-related convergence failures (c.f., Jonckere & Rosseel, 2025; Rosseel & Loh, 2024)
- ▶ Guaranteed convergence
- ▶ Easy case-wise statistics
- ▶ Convergence despite model unidentifiability...?
- ▶ Handles both components and factors (c.f., Schamberger et al., 2025)

Cons of IGSCA Trees:

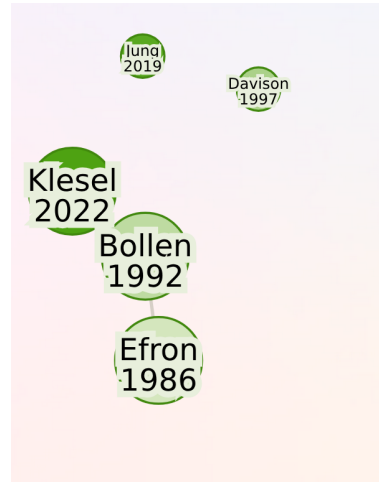
- ▶ Bootstrapped measures of uncertainty (Tests/SEs)
- ▶ Converges to local minima
- ▶ Potentially over-parameterized/un-identified ($\approx$ NNs)
- ▶ Very computationally intensive

1. IGSCA implementation: Correct?
2. Best multigroup hypothesis testing and bootstrapping?
3. Constrained least-squares optimization
4. SEM trees vs. IGSCA trees

INFERRING CORRECTNESS BY REPLICATING SIMULATION RESULTS



MULTIGROUP HYPOTHESIS TESTING AND BOOTSTRAPPING PROCEDURES



WHAT IS THE POINT? WHY SHOULD I CARE?

1. First IGSCA Trees in R + Fast
2. First investigation of MGA for IGSCA
3. First SEM Trees vs. IGSCA Trees



Manuscript 3: Machine Learning Structural Equation Modeling and Falsificatory Data Analysis – A Tutorial



TABLE OF CONTENTS

BUT WHAT DOES ML-SEM HAVE TO DO WITH FDA?

EXAMPLE 1: DATA QUALITY CHECK

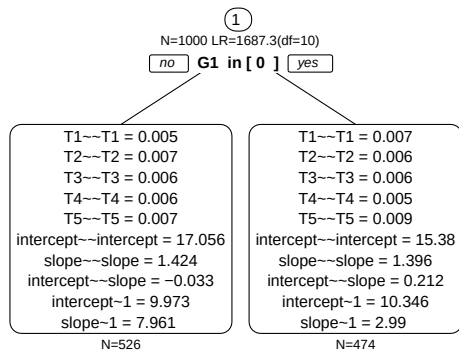
EXAMPLE 2: BASIC EMOTION THEORY

EXAMPLE 3: FALSIFICATION TESTS

WHAT IS THE POINT? WHY SHOULD I CARE?

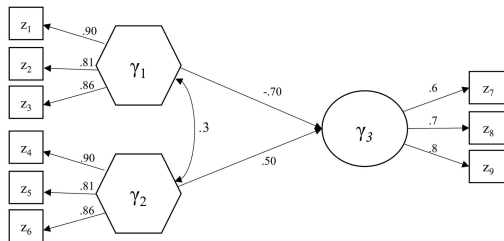
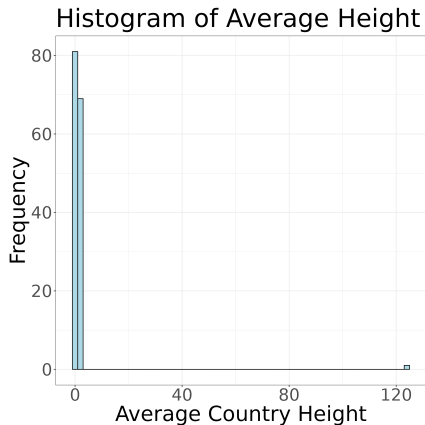
Central Tension:

- ▶ SEM Trees may overfit: semblance of meaning out of meaningless noise
- ▶ However, null hypothesis is quasi-always false, *there is always a difference* (Gelman & Carlin, 2014)
- ▶ Side-step: Severely test theoretically impossible differences using SEM Trees
 - ▶ Anti-predictors
 - ▶ Zone of impossibility



Let's put everything together!

HOW CAN I USE ML-SEM & FDA TO ASSESS DATA QUALITY?



ML-SEM + FDA = ?
○○

EX1:QC
○○●

EX2:BET
○○○

EX3:FT
○○

EXPECTED OUTCOMES
○○

+

WHAT ARE THE ASSERTIONS OF BASIC EMOTION THEORY?

'Basic emotions' are said to be genetically inherited phenotypic characteristics that are specific adaptations with fixed physical forms and fixed functions across time. For example, instances of FEAR are thought to manifest with a small degree of variation around a universal pattern of physiological changes, expressive behaviors, and so on (see van Heijst et al., 2025, p. 377). This pattern is hypothesized to have evolved as an adaptation to navigate the recurrent fitness challenge of encountering predatory threat. Its cause is reduced to a single, innate (hardwired) neural circuit that does not change or function differently across time, context, or individual. (Barrett et al., 2025, p.403).

HOW CAN I USE ML-SEM & FDA TO TEST BASIC EMOTION THEORY?

See whiteboard

Anti-predictor: Person's culture and context of stimuli presentation

ML-SEM + FDA = ?
○○

EX1:QC
○○

EX2:BET
○○●

EX3:FT
○○

EXPECTED OUTCOMES
○○

+

HOW CAN I USE ML-SEM + FDA TO VERIFY SCIENTIFIC ASSUMPTIONS?

- ▶ The hippocampal + MTL brain regions $\rightsquigarrow$ Episodic Memory
- ▶ Episodic memory concerns memory of events that happened in the past
- ▶ Let's study the neural instantiation of time by comparing memory of events 'with' and 'without' temporal ordering
- ▶ We begin with a behavioral study to validate our paradigm before commencing costly neuroimaging

A **falsification test** is when we attempt to falsify “empirically testable implications of the identifying assumptions” (Hartman, 2021, p.505) behind a causal estimate; failures to (statistically) falsify these assumptions should then raise confidence that a causal estimate is acceptable.

Falsification test using GLMM Trees

```
glmer(hit ~ strategy | (1 | ID) | counterbalance)
```

ML-SEM + FDA = ?
○○

EX1:QC
○○

EX2:BET
○○○

EX3:FT
○●

EXPECTED OUTCOMES
○○

+

WHAT IS THE POINT? WHY SHOULD I CARE?

Current and concrete demonstration of the productive synergy of FDA methodology and ML-SEM methods.

From theory of ML-SEM & FDA to practice of ML-SEM & FDA

ML-SEM + FDA = ?
○○

EX1:QC
○○

EX2:BET
○○○

EX3:FT
○○

EXPECTED OUTCOMES
○○●

+

General Conclusion & Big Picture

REPLY TO PAUL E. MEEHL

Risky tests of specific predictions made by theories is an important idea, but is beset by too many challenging problems.

Assertion: the zone of possibility is far larger than the zone of impossibility.

Instead, let's do risky tests of specific 'anti-predictions' made by theories.

WHAT ARE THE GENERAL EXPECTED OUTCOMES?



WHAT COULD BE THE CENTRAL DISEASE?



Assertion: You cannot replace a very poor theory with 'no theory'.

HOW MIGHT THIS CENTRAL DISEASE BE CURED?



THANK YOU!

Special thanks to: Drs. Ji Yeh Choi, David B. Flora, R. Phil Chalmers, Ellen Bialystok, Andreas Brandmaier, Rob Cribbie, John Fox, Michael Friendly, Heungsun Hwang, Georges Monette, Morris Moscovitch, W. Dale Stevens, Florian Schuberth & Cathy Xijuan Zhang.

Thanks does not imply their endorsement of my work! Contact: emstruonger@gmail.com

Appendix

OBJECTIVE FUNCTIONS

$$\begin{aligned}\max [FIT] &= 1 - \frac{SS [\mathbf{\Psi} - \mathbf{\Gamma A} - \mathbf{V}]}{SS [\mathbf{\Psi}]} \\ &= \min [SS (\mathbf{\Psi} - \mathbf{\Gamma A} - \mathbf{V})] \\ \hat{\mathbf{\Gamma}}_{LV} &= (Z - UD) M H^T (H H^T)^{-1} \\ \text{vec}(\mathbf{A})_{LV} &= (\mathbf{\Phi}^T \mathbf{\Phi})^{-1} \mathbf{\Phi}^T \text{vec}(\mathbf{\Psi} - \mathbf{S})\end{aligned}$$

$$F_{ML} = \ln |\Sigma| + \text{Tr}[\mathbf{S}\Sigma^{-1}] - \ln |\mathbf{S}| - (p + q)$$

WHAT'S THE (MY) PROBLEM WITH TCE?

- ▶ It seems that **Theory of Constructed Emotions'** (TCE) conditions for falsification are completely different from BET
- ▶ **Basic Emotion Theory** (BET) is almost too easy to falsify
- ▶ What TCE and BET each consider to be signal vs noise in data, is completely different

Assertion: We can only meaningfully falsify BET and make progress because TCE exists. Otherwise, we're forced to hold onto BET because **you cannot replace a 'bad' theory with no theory.**

NOMENCLATURE I

Γ Matrix of construct scores

S Measurement Error

V Stacked Matrix of S and 0s

H $H = A - [0 \mid I_p]$

M $M = [I_j \mid 0]$

REFERENCES I



Barrett, L. F. (2017). *How Emotions Are Made: The Secret Life of the Brain*. Mariner Books.



Barrett, L. F., Atzil, S., Bliss-Moreau, E., Chanes, L., Gendron, M., Hoemann, K., Katsumi, Y., Kleckner, I. R., Lindquist, K. A., Quigley, K. S., Satpute, A. B., Sennesh, E., Shaffer, C., Theriault, J. E., Tugade, M., & Westlin, C. (2025). The Theory of Constructed Emotion: More Than a Feeling. *Perspectives on Psychological Science*, 20(3), 392–420. <https://doi.org/10.1177/17456916251319045>



Cinelli, C., Forney, A., & Pearl, J. (2024). A Crash Course in Good and Bad Controls. *Sociological Methods & Research*, 53(3), 1071–1104. <https://doi.org/10.1177/00491241221099552>



Furukawa, T. A., Katherine Shear, M., Barlow, D. H., Gorman, J. M., Woods, S. W., Money, R., Etschel, E., Engel, R. R., & Leucht, S. (2009). Evidence-based guidelines for interpretation of the Panic Disorder Severity Scale. *Depression and Anxiety*, 26(10), 922–929. <https://doi.org/10.1002/da.20532>



Gelman, A., & Carlin, J. (2014). Beyond Power Calculations: Assessing Type S (Sign) and Type M (Magnitude) Errors. *Perspectives on Psychological Science*, 9(6), 641–651. <https://doi.org/10.1177/1745691614551642>



Gelman, A., Vehtari, A., Simpson, D., Margossian, C. C., Carpenter, B., Yao, Y., Kennedy, L., Gabry, J., Bürkner, P.-C., & Modrák, M. (2020). Bayesian Workflow. <https://doi.org/10.48550/arXiv.2011.01808>



Google. (2025a, September 28). *Gen Art QM Sep 29 - 2025*. Gemini 2.5 Flash Image. Retrieved September 29, 2025, from https://aistudio.google.com/app/prompts?state=%7B%22ids%22:%5B%221Lc4o_70jqmZhUTgW_jbsEHe79jTN8ew0%22%5D,%22action%22:%22open%22,%22userId%22:%2211616366773954702139%22,%22resourceKeys%22:%7B%7D%7D&usp=sharing

REFERENCES II



Google. (2025b, September 23). *Gondor Tree & Lucifer Painting*. Gemini 2.5 Flash Image. Retrieved September 23, 2025, from <https://aistudio.google.com/app/prompts?state=%7B%22ids%22:%5B%221AO-d0iq6FATR040PFun6RxH3YLYHaD2s%22%5D,%22action%22:%22open%22,%22userId%22:%22111616366773954702139%22,%22resourceKeys%22:%7B%7D%7D&usp=sharing>



Hartman, E. (2021). Equivalence Testing for Regression Discontinuity Designs. *Political Analysis*, 29(4), 505–521. <https://doi.org/10.1017/pan.2020.43>



Hong, M., Jacobucci, R., & Lubke, G. (2020). Deductive data mining. *Psychological Methods*, 25(6), 691–707. <https://doi.org/10.1037/met0000252> (US).



Jonckere, J. D., & Rosseel, Y. (2025). A note on using random starting values in small sample SEM. *Behavior Research Methods*, 57(1), 57. <https://doi.org/10.3758/s13428-024-02543-9>



Masur, P. K., & Scharkow, M. (2020). Specr: Conducting and visualizing specification curve analyses (version 1.0.0). <https://CRAN.R-project.org/package=specr>



McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan* (2nd ed.). Taylor and Francis, CRC Press.



Nadel, L., & Moscovitch, M. (1997). Memory consolidation, retrograde amnesia and the hippocampal complex. *Current Opinion in Neurobiology*, 7(2), 217–227. [https://doi.org/10.1016/s0959-4388\(97\)80010-4](https://doi.org/10.1016/s0959-4388(97)80010-4)



Pearl, J., & Mackenzie, D. (2018). *The book of why: The new science of cause and effect*. Basic Books.

REFERENCES III



Robinaugh, D. J., Haslbeck, J. M. B., Waldorp, L. J., Kossakowski, J. J., Fried, E. I., Millner, A. J., McNally, R. J., Ryan, O., de Ron, J., van der Maas, H. L. J., van Nes, E. H., Scheffer, M., Kendler, K. S., & Borsboom, D. (2024). Advancing the network theory of mental disorders: A computational model of panic disorder. *Psychological Review*, 131(6), 1482–1508. <https://doi.org/10.1037/rev0000515>



Rosseel, Y., & Loh, W. W. (2024). A structural after measurement approach to structural equation modeling. *Psychological Methods*, 29(3), 561–588. <https://doi.org/10.1037/met0000503>



Schamberger, T., Schuberth, F., Henseler, J., & Rosseel, Y. (2025, August 8). *Structural Equation Modeling with Latent Variables and Composites*. arXiv: 2508.06112 [stat]. <https://doi.org/10.48550/arXiv.2508.06112>



Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2020). Specification curve analysis. *Nature Human Behaviour*, 4(11), 1208–1214. <https://doi.org/10.1038/s41562-020-0912-z>



Tosh, C., Greengard, P., Goodrich, B., Gelman, A., Vehtari, A., & Hsu, D. (2025). The Piranha Problem: Large Effects Swimming in a Small Pond. *Notices of the American Mathematical Society*, 72(01), 1. <https://doi.org/10.1090/noti3044>



Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, 185(4157), 1124–1131. <https://doi.org/10.1126/science.185.4157.1124>



Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 90(4), 293–315. <https://doi.org/10.1037/0033-295X.90.4.293>

REFERENCES IV



Wysocki, A. C., Lawson, K. M., & Rhemtulla, M. (2022). Statistical Control Requires Causal Justification. *Advances in Methods and Practices in Psychological Science*, 5(2), 25152459221095823. <https://doi.org/10.1177/25152459221095823>