A Holistic Bayesian Approach to Addressing Measurement Reactivity with a Planned Missing Data Design

Mark Himmelstein & David Budescu

Fordham University

Presented at the Modern Modeling Methods Conference

University of Connecticut

June 27, 2023

Contact: mhimmelstein@fordham.edu

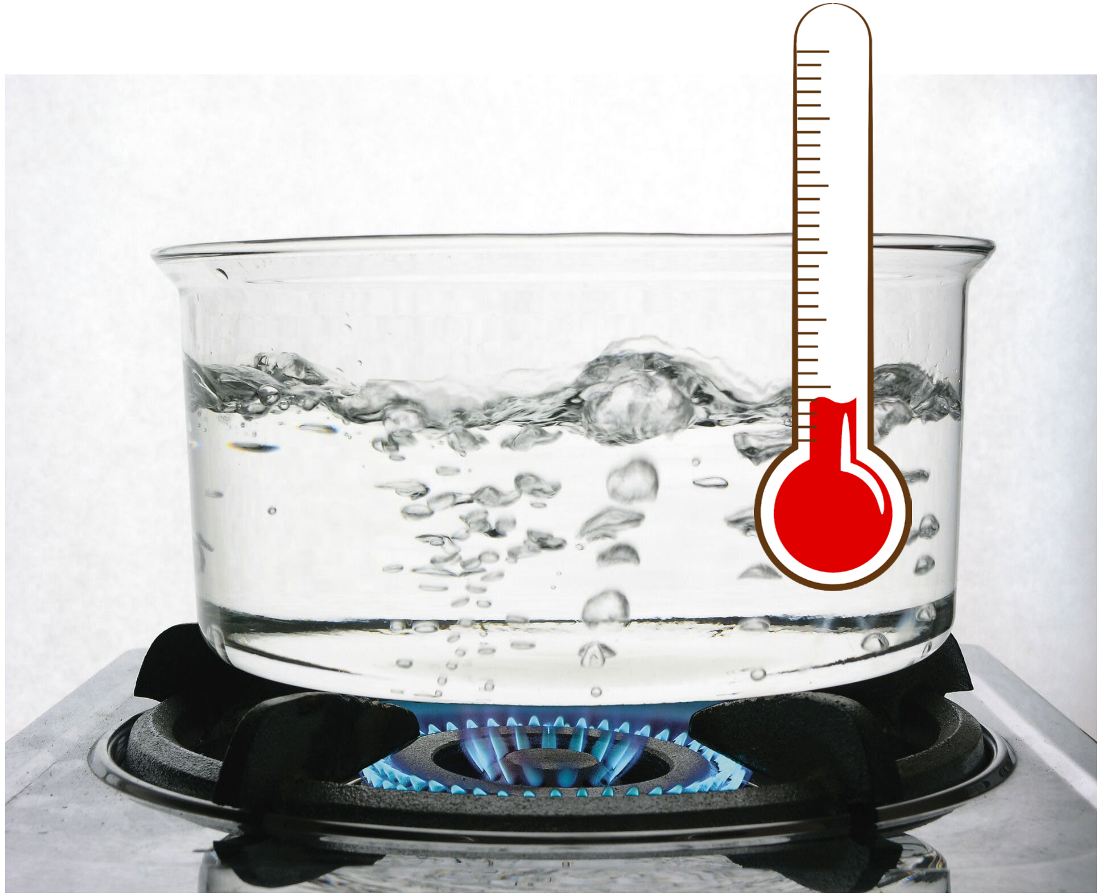
Outline of Talk

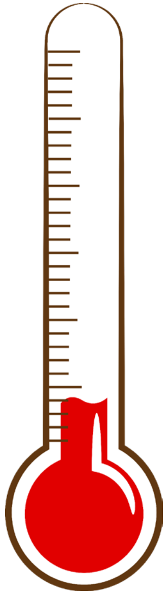
1. Define **measurement reactivity** and show how it confounds research with **pre-test post-test designs**
2. Redefine measurement reactivity as a **missing data problem**
3. Show example in the literature for using **missing data imputation** to cure a **confounding effect of measurement reactivity**
4. Describe a classic **advice taking experimental design** as a simple example of a **pre-test post-test design**
5. Demonstrate how **advice changes people's minds** when their **prior beliefs are not directly measured** using fully Bayesian **joint imputation and measurement model**

The Observer Effect

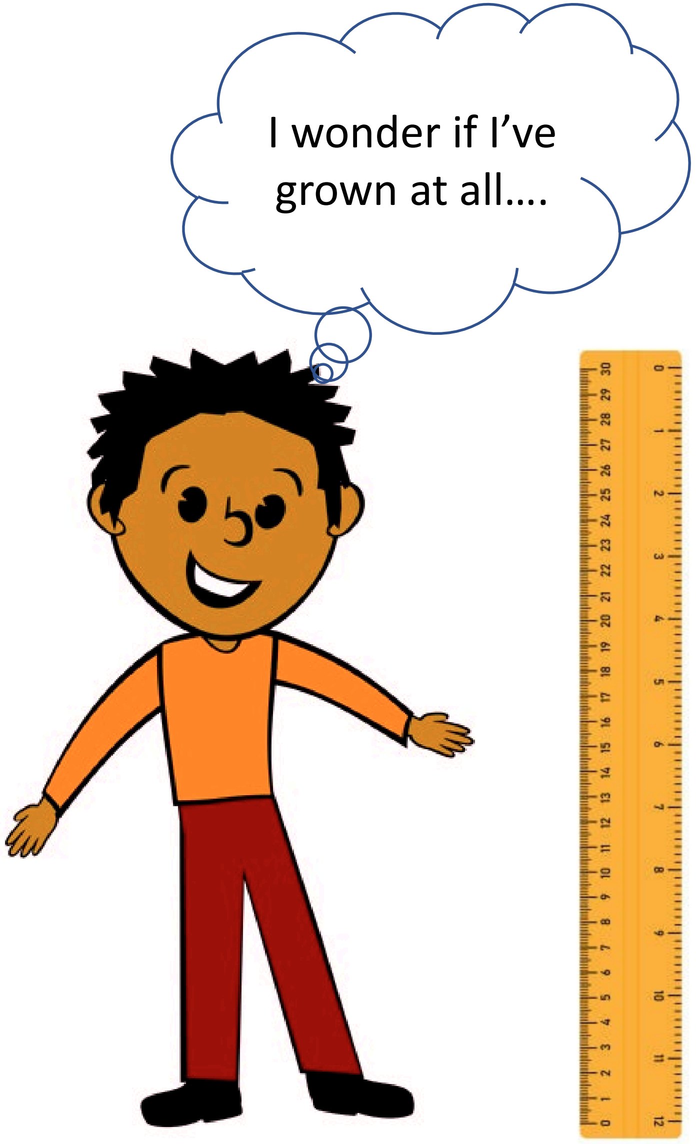
It is impossible to measure something without changing it

250°
249.999°





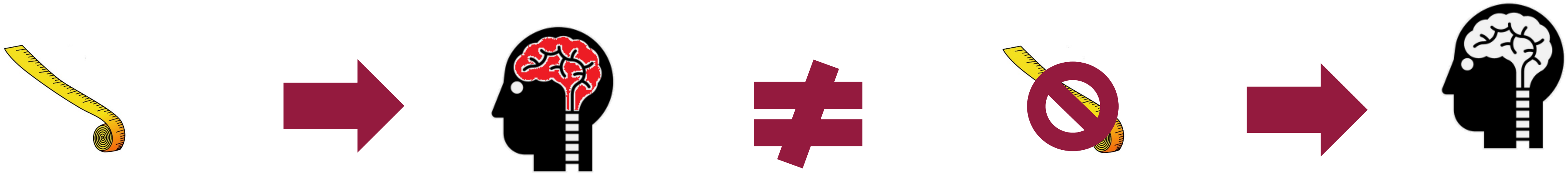
Measurement Reactivity



When Does Measurement Reactivity Matter?

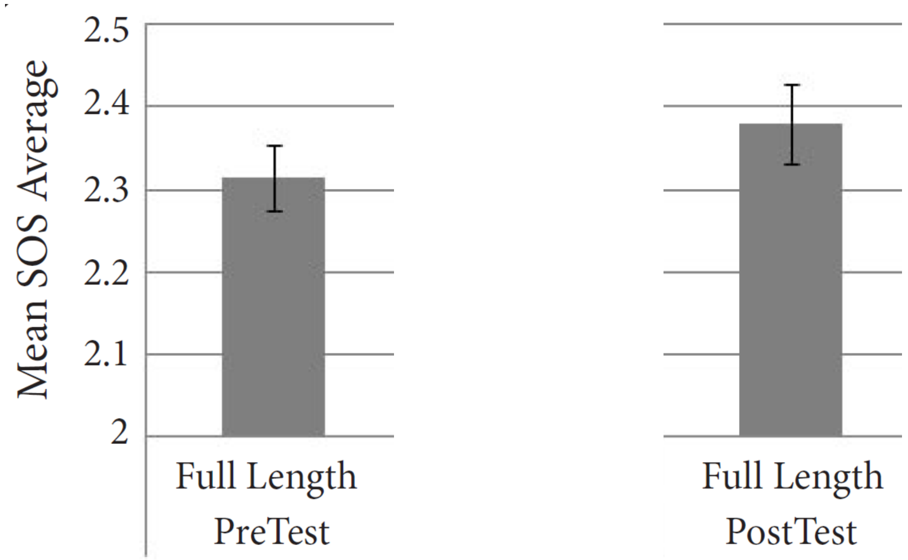
We often wish to understand the change that occurs as a result of a **treatment** or **intervention**

But existence of a pretest can change the effects of interventions and **threaten generalizability**



Treatment Effects in Control Groups

- In RCTs with pre-tests and post-tests, results often show treatment-like effects even in control groups
- Such results reflect learning as a result of pre-test exposure
- Harel et al. (2011) showed increases in scores on a measure of attitudes toward suicide among control participants



Missing Data Mechanisms

Terms

- **missing**: Patterns of missing data
- **data**: Values of the actual data
- **x:** Set of completely observed predictors correlated with incomplete data *and* patterns of missingness

Missing Completely At Random: Patterns of missingness are independent of the data that is missing

- P(missing|data, x) = P(missing)

Missing at Random: Patterns of missingness are *conditionally* independent of the data that is missing

- P(missing|data, x) = P(missing | x)

Imputation Possible

Missing Not at Random: Patterns of missingness are not independent of the data that is missing

- P(missing|data, x) ≠ P(missing | x)

Imputation Not Possible

Reframing Measurement Reactivity

Missing data mechanisms deal with whether ***missingness depends on data***

Measurement reactivity is a ***counterfactual*** question about whether ***data depends on missingness***

Measurement reactivity: $P(\text{data} | \text{missing}) \neq P(\text{data})$

Modeling Solution

- Find domains where measurement reactivity is a problem
  - $P(data|missing) \neq P(data)$
- Ensure assumptions for imputation are met
  - MCAR: $P(missing|data) = P(missing)$, or at least
  - MAR: $P(missing|data, x) = P(missing \mid x)$
- Impute missing values
- Compare imputed missing values to observed values to understand patterns in data in the absence of measurement reactivity

Planned Missing Data Design

Sometimes collectable data is left missing by design (Graham et al., 2006; Rhemtulla & Hancock 2016). Typically done to reduce

- Instrument length
- Participant burden
- Attrition

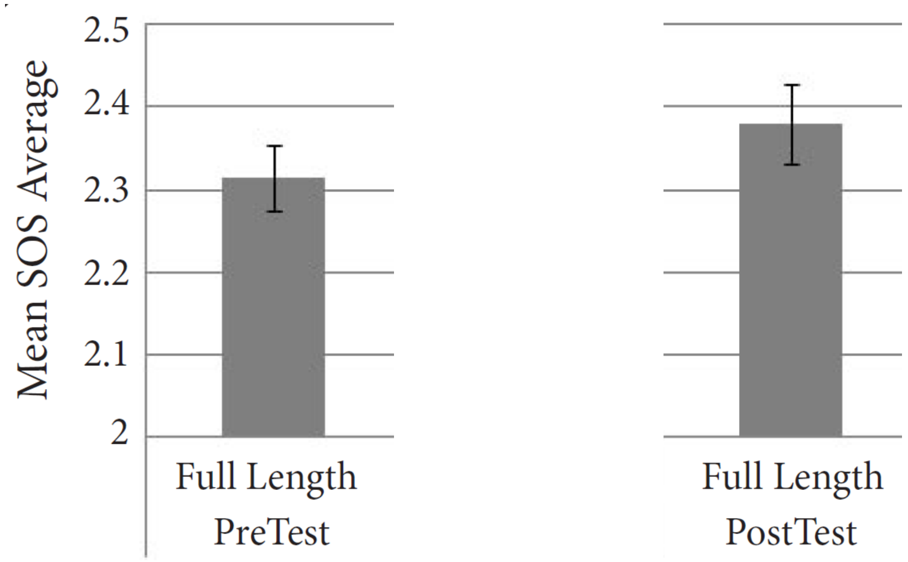
For measurement reactivity research, we do something similar

- Experimentally randomize missing data
- Impute
- Draw causal inference

Treatment Effects in Control Groups

- In RCTs with pre-tests and post-tests, results often show treatment-like effects even in control groups
- Such results reflect learning as a result of pre-test exposure
- Harel et al. (2011) showed increases in scores on a measure of attitudes toward suicide among control participants

- However, when truncated pretests were used with multiple imputation for the missing data, no treatment effects were found in control group



Multiple Imputation vs. Fully Bayesian Approach

Multiple Imputation

- Fit imputation model
- Extract subset of posterior draws to represent uncertainty in missing values
- Fit measurement model on each draw
- Combine/average results of measurement models

Fully Bayesian

- Fit imputation and measurement model with joint posterior
- Treat missing values as parameters / latent variables in measurement model
- Utilize full posterior uncertainty of missing values
- Only need to fit one model

Fully Bayesian Imputation in Stan

https://mc-stan.org/docs/stan-users-guide/missing-data.html

3.1 Missing data

Stan treats variables declared in the `data` and `transformed data` blocks as known and the variables in the `parameters` block as unknown.

An example involving missing normal observations could be coded as follows.[10]

```
data {
  int<lower=0> N_obs;
  int<lower=0> N_mis;
  array[N_obs] real y_obs;
}
parameters {
  real mu;
  real<lower=0> sigma;
  array[N_mis] real y_mis;
}
model {
  y_obs ~ normal(mu, sigma);
  y_mis ~ normal(mu, sigma);
}
```

Observed data

Missing data

The number of observed and missing data points are coded as data with non-negative integer variables `N_obs` and `N_mis`. The observed data are provided as an array data variable `y_obs`. The missing data are coded as an array parameter, `y_mis`. The ordinary parameters being estimated, the location `mu` and scale `sigma`, are also coded as parameters. The model is vectorized on the observed and missing data; combining them in this case would be less efficient because the data observations would be promoted and have needless derivatives calculated.

Advice Taking: Judge-Advisor Systems (JAS)

I wonder if the Mets are likely to make the playoffs

Agent or “judge” tasked with forming an objective belief

They Stink! Maybe a 10% chance!

Consults an “advisor” for new information

FANGRAPHS

Team	Make Playoffs
Mets	42.2%

Can consult many advisors, who need not be people

Who am I kidding, I’ll go with 25%

Judge makes decision

Studying Belief Change

To study belief-change as a result of advice, researchers will typically ask a judge what they believe **before** and **after** getting advice

Pre-test:
Initial judgment (J_1)

I think maybe 75%!

Treatment:
Advice (J_a)

They Stink! Maybe a 10% chance!

Post-test:
Final judgment (J_2)

Maybe Stephen A is right, I'll go with 25% instead

Weight of Advice (WOA)

$$WOA = \frac{J_2 - J_1}{J_a - J_1} = \frac{25\% - 75\%}{10\% - 75\%} = .77$$

Implies

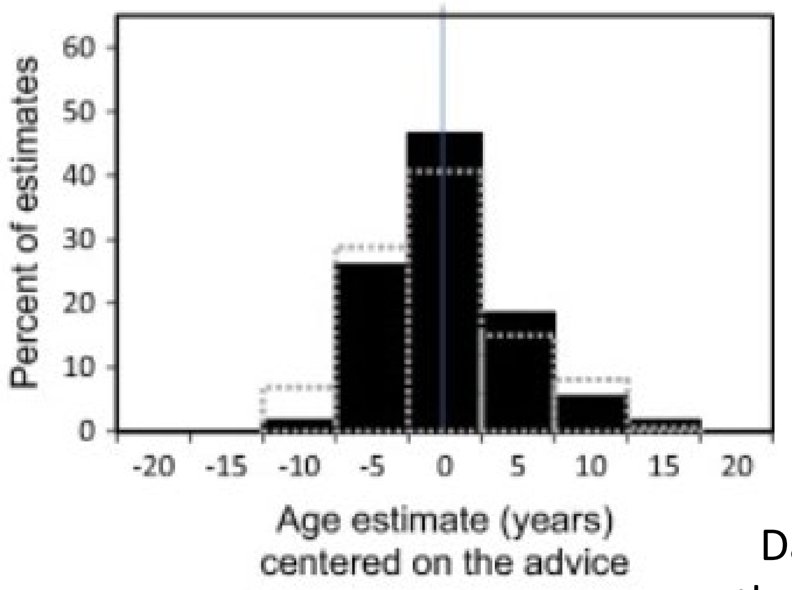
$$J_2 = WOA(J_a) + (1 - WOA)(J_1) = 25\% = .77(10\%) + .23(75\%)$$

What happens if the pre-test is removed?

Pushing Away vs Staying Close

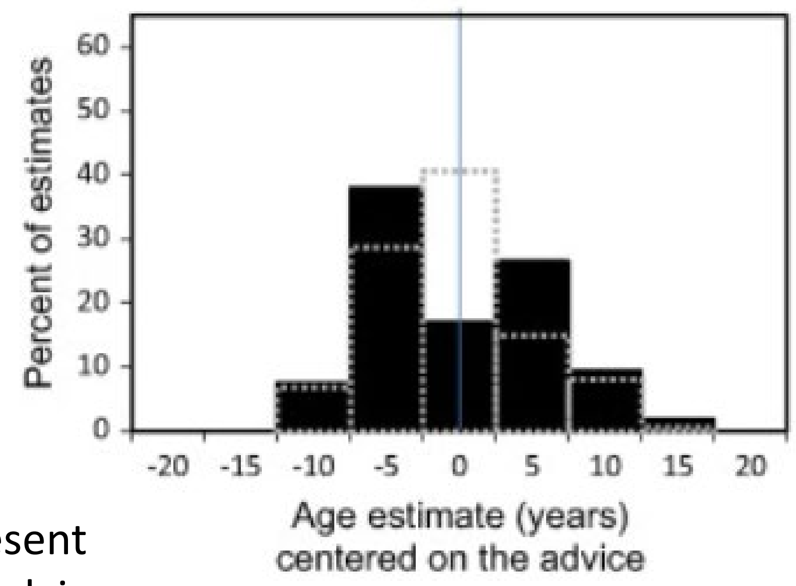
(Rader et al., 2015)

Untruncated JAS: Final Judgment (J_2) when initial judgment (J_1) WAS elicited



No “push away” effect

Truncated JAS: Final Judgment (J_2) when initial judgment (J_1) WAS NOT elicited



“Push away” effect

Dashed bars represent the distribution of advice

Difference implies P(data|missing) ≠ P(data)

Solution: Planned Missing Data Experimental Design

*Un*truncated Condition

Type	Item #	J_1	J_2
mis	1	Observed	Observed
mis	2	Observed	Observed
mis	⋮	Observed	Observed
mis	*m*	Observed	Observed
obs	1	Observed	Observed
obs	2	Observed	Observed
obs	⋮	Observed	Observed
obs	*o*	Observed	Observed

Partially *Truncated* Condition

Type	Item #	J_1	J_2
mis	1	**MISSING!**	Observed
mis	2	**MISSING!**	Observed
mis	⋮	**MISSING!**	Observed
mis	*m*	**MISSING!**	Observed
obs	1	Observed	Observed
obs	2	Observed	Observed
obs	⋮	Observed	Observed
obs	*o*	Observed	Observed

Imputation Model

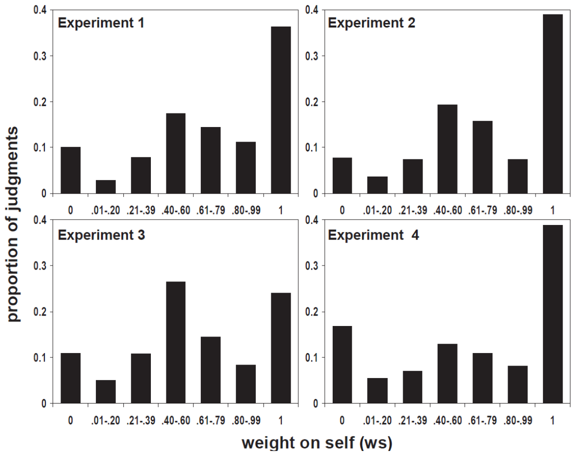
$J_{1,mis,un} = \alpha' z_{un} + J_{1,obs,un} B + \epsilon_{un}$

$\delta_{mis,tr} = \alpha' z_{tr} + J_{1,obs,tr} B$

$\hat{J}_{1,mis,tr} \sim MVN(\delta_{mis,tr}, \hat{\Sigma}_{\epsilon_{un}})$

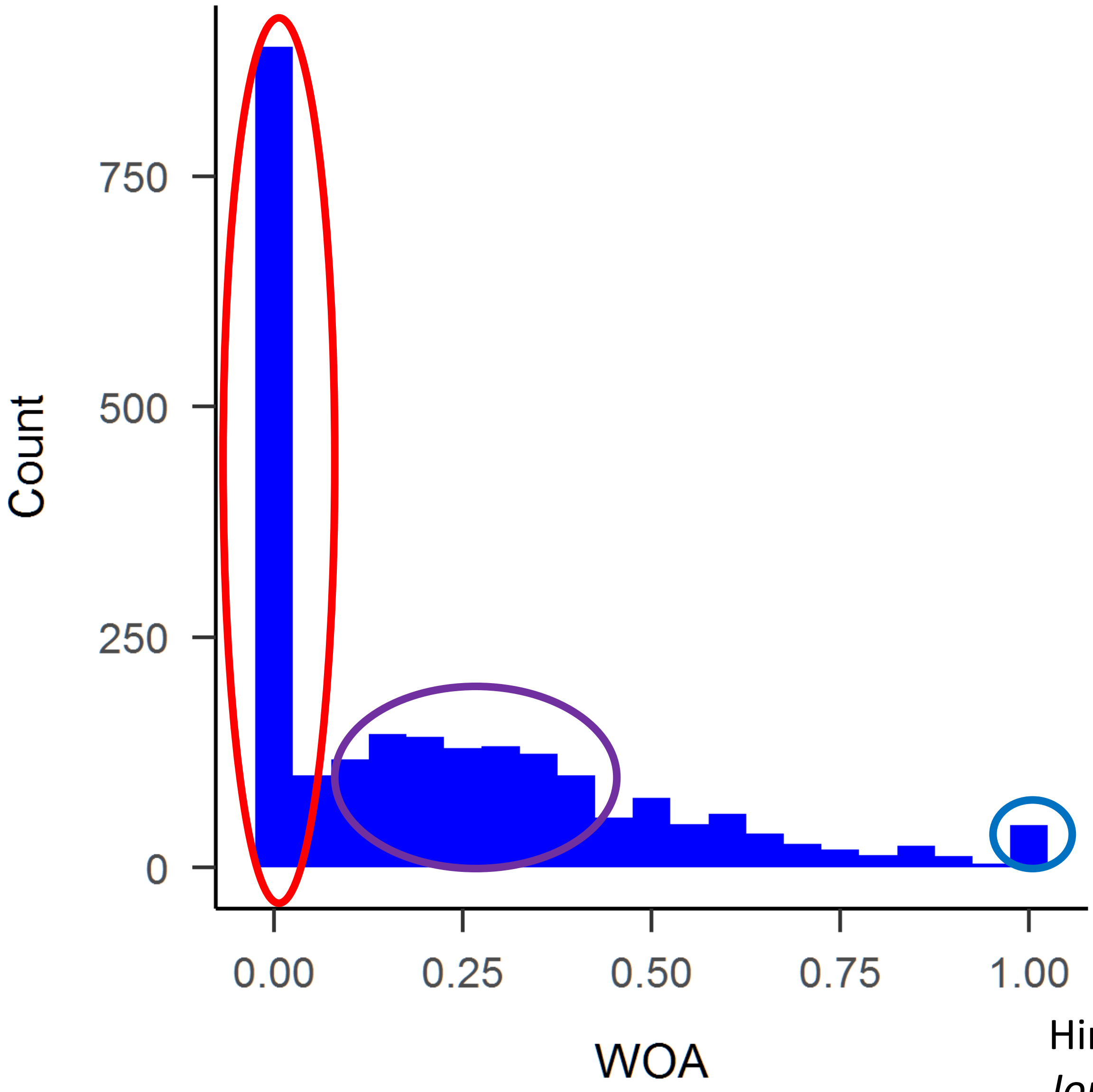
The Empirical Distribution of Untruncated WOA

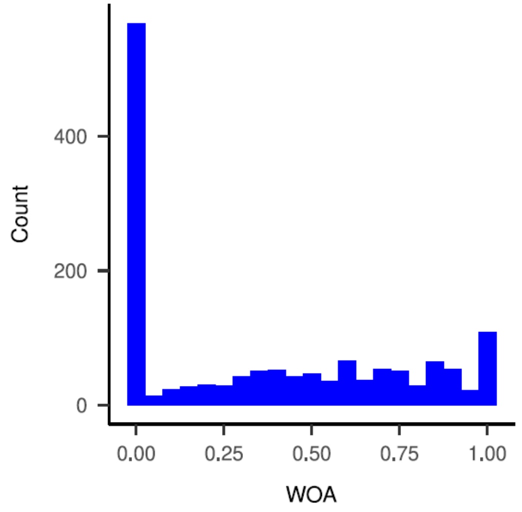
Three
Distinct
Modes

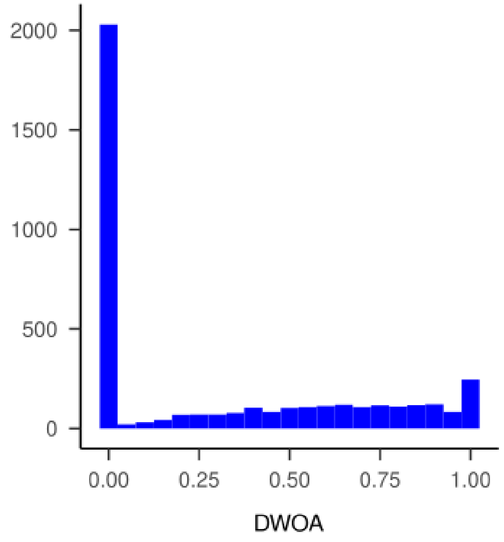


(Soll & Larrick, 2009)

Note: *ws* = 1 - WOA







Himmelstein, 2022
Journal of Mathematical Psychology

Untruncated JAS: Dual Hurdle Model

- Treats WOA $\in [0,1]$ as response variable
- Mixture between
  - **Categorical model**: Decline (WOA = 0), adopt (WOA = 1), or compromise (0 < WOA < 1)
  - **Continuous model**: averaging judgment conditional on compromise

Truncated JAS: Hurdled Mixture Model

- Treats $J_2 \in (-\infty, \infty)$ as response variable
- Adopt ($J_2 = J_a$)remains categorical submodel
- Imputed $\hat{J}_1$ contain measurement error, meaning $\hat{J}_1 = J_2$ exactly will not occur naturally
- Replace discrete *decline* observations with a Gaussian component that incorporates measurement error

Measurement Model: Hurdled Mixture

$$J_2 \begin{cases} = J_{a,mis,tr} \text{ with } P_{adopt} = \eta \\ = \hat{J}_1 \text{ with } P_{decline} = \tau(1-\eta) \\ = \hat{C} \text{ with } P_{compr} = (1-\tau)(1-\eta) \end{cases}$$

$$\hat{\boldsymbol{J}}_1 \sim MVN(\boldsymbol{\delta}_{mis,tr}, \widehat{\boldsymbol{\Sigma}}_{\epsilon_{un}})$$

$$\hat{\mu} = \omega J_a + (1-\omega)\hat{J}_1$$

$$\hat{C} \sim N(\hat{\mu}, \hat{\sigma}_{compr})$$

Simulation Study

Simulated population correlation matrices $\boldsymbol{\Sigma}_{\boldsymbol{J_1}}$ for *K* items. Each iteration varied:

- γ*:* Proportion of variance accounted by first eigenvalue of $\boldsymbol{\Sigma}_{\boldsymbol{J_1}}$
- *K:* Total number of items
- *N:* Total number of subjects
- *S:* Proportion of items in partially truncated condition
- *D*: Rate of declining advice
  - **Hypothesis 1:** ~47.5%, people anchor on prior belief, like untruncated JAS (big mode at $J_1 = J_2$)
  - **Hypotheses 2:** ~2.5%, people do not anchor on prior belief, unlike untruncated JAS (no big mode at $J_1 = J_2$)

Simulation Study

- **Step 1:** Simulate *N* judges and their prior belief ($\boldsymbol{J_1}$) values for each of the *K* items based on $\boldsymbol{\Sigma}_{\boldsymbol{J_1}}$
- **Step 2:** Simulate advice $\boldsymbol{J_a}$ for each item (based on $\boldsymbol{\Sigma}_{\boldsymbol{J_1}}$ with a small amount of residual error)
- **Step 3:** simulate for each judge a parameter that defines
  - Stage 1 tendencies (discrete decision, η_j)
  - Stage 2 tendencies (continuous averaging judgment, ω_j)

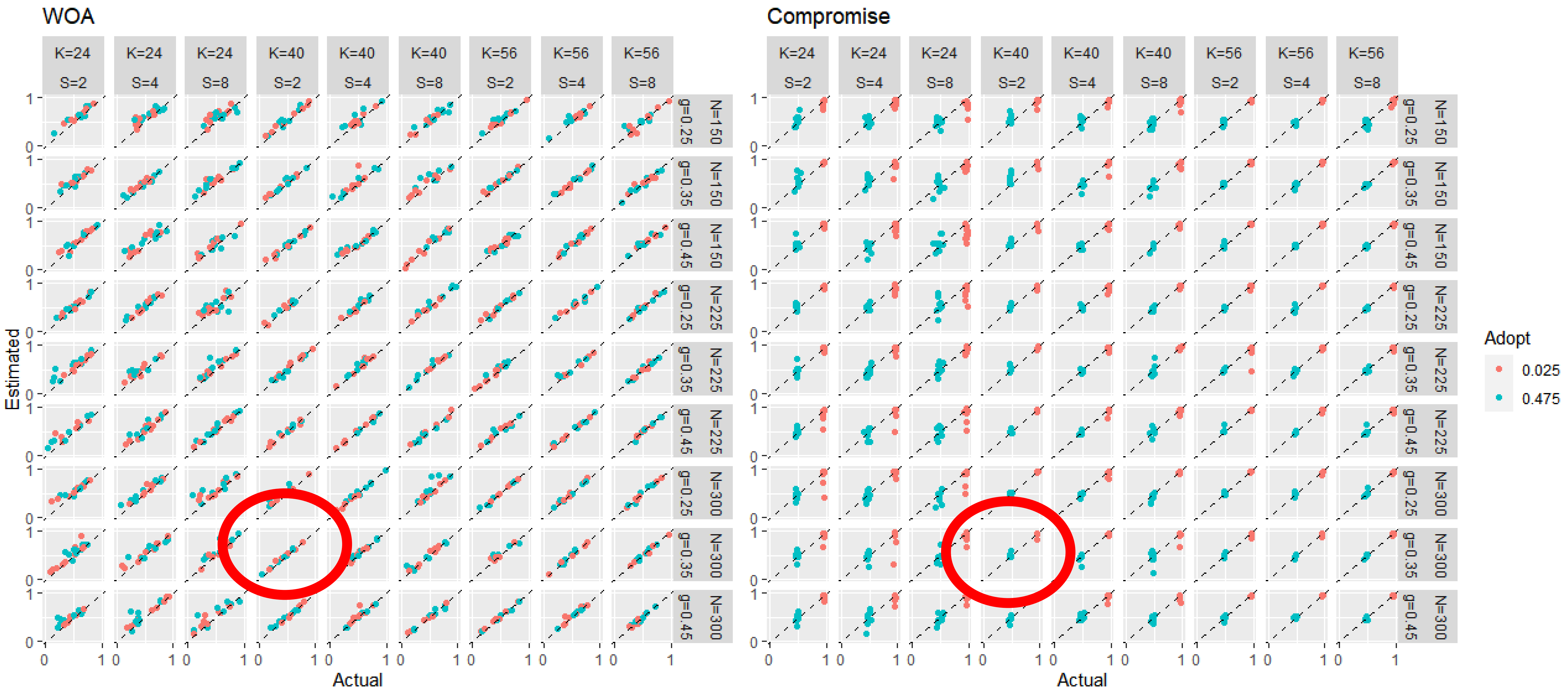
Simulation Study

- **Step 4:** For each combination *N* and *K* simulate a realized outcome for discrete decision ($\eta_{j,k}$) and averaging judgment ($\omega_{j,k}$)
- **Step 5:** For each judge and item, define J_2 as follows
  - If discrete decision = decline, $J_2 = J_1$
  - If discrete decision = adopt, $J_2 = J_a$
  - If discrete decision = compromise, $J_2 = \omega_{j,k} J_a + (1 - \omega_{j,k}) J_1$
- **Step 6:** Randomly select $\frac{N}{3}$ simulated judges and $\frac{K}{S}$ simulated items and delete J_1 in all cases

Study 1: Simulated Schultze (2015) data

- **Step 7:** Use observed values from remaining $\mathrm{N} - \frac{N}{3}$ judges to estimate imputation model that uses remaining $\mathrm{K} - \frac{K}{S}$ untruncated items to predict $\frac{K}{S}$ items that were artificially truncated
- **Step 8:** Impute J_1 values that were deleted in **Step 6** based on imputation model estimated in **Step 7**
- **Step 9:** estimate hurdled mixture model using values imputed in **Step 8** and compare the estimated $\hat{\eta}$ and $\widehat{\omega}$ to actual simulated η and ω

Results


WOA
Compromise
K=24 S=2
K=24 S=4
K=24 S=8
K=40 S=2
K=40 S=4
K=40 S=8
K=56 S=2
K=56 S=4
K=56 S=8
g=0.25 N=150
g=0.35 N=150
g=0.45 N=150
g=0.25 N=225
g=0.35 N=225
g=0.45 N=225
g=0.25 N=300
g=0.35 N=300
g=0.45 N=300
Estimated
Actual
Adopt
0.025
0.475

A Real Truncated JAS

- Schultze et al. (2015) design was a JAS in which subjects estimated the number of calories in different foods.
- Took images from 40 different **Blue Apron** recipes to use in a JAS in three conditions
  - Untruncated (N = 196)
  - Partially Truncated (N = 148)
  - Fully truncated (N = 52, control condition to compare against partially truncated condition)

Untruncated Item Example

0 150 300 450 600 750 900 1050 1200 1350 1500

How many calories (kcal) are in this dish? 0

Your **advisor** made the following estimate: **450 Calories**

0 150 300 450 600 750 900 1050 1200 1350 1500

How many calories (kcal) are in this dish? 633

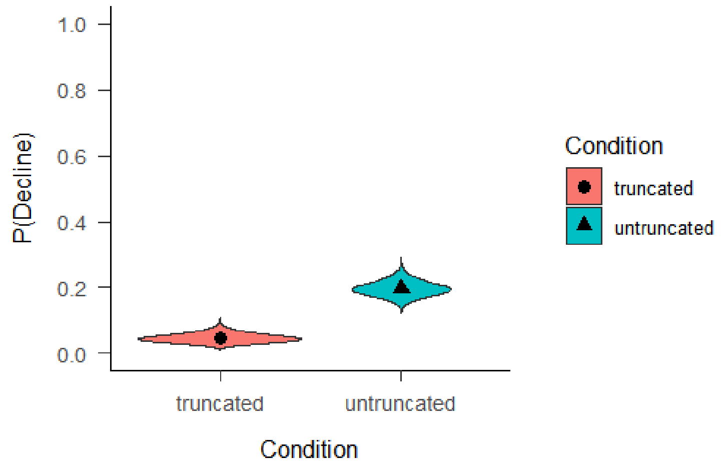
Truncated Item Example

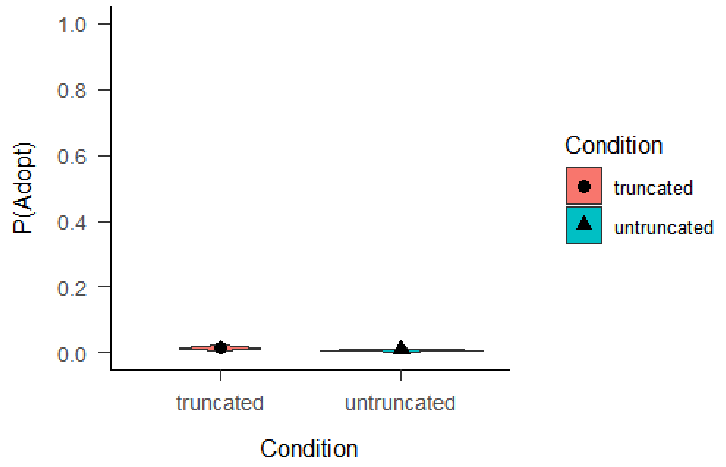
Your **advisor** made the following estimate: **600 Calories**

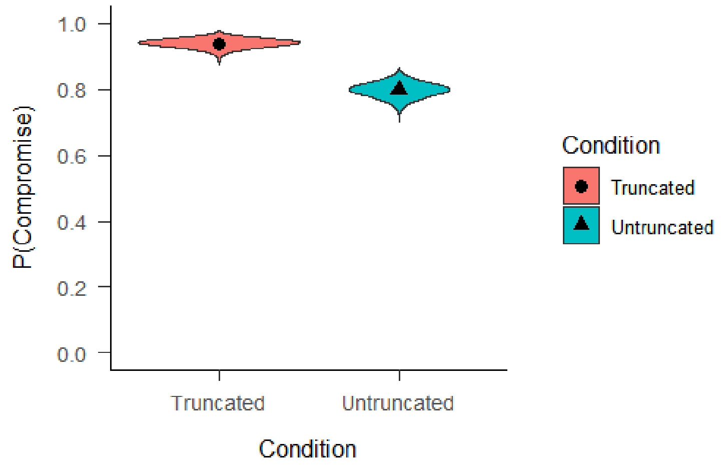
0 150 300 450 600 750 900 1050 1200 1350 1500

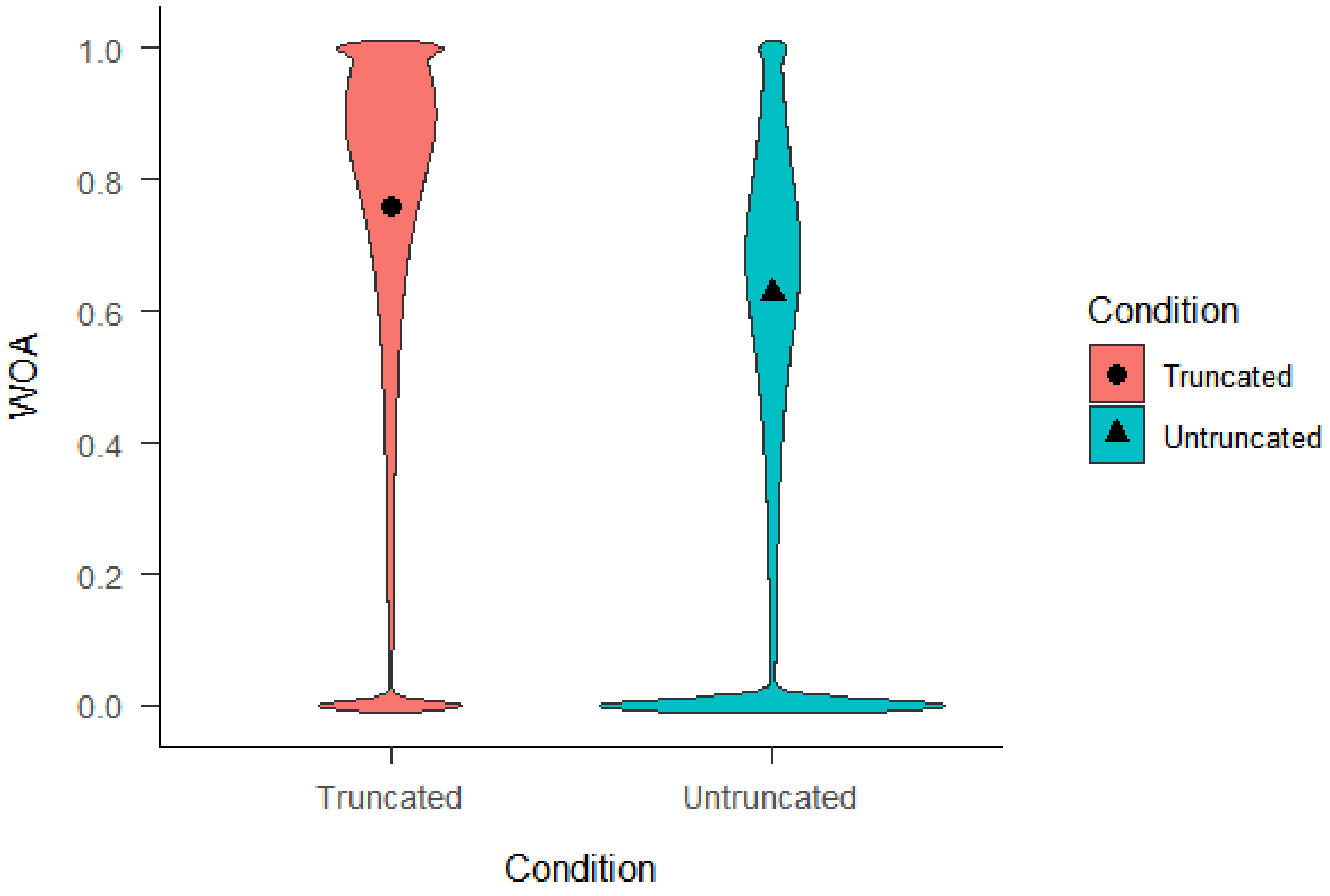
How many calories (kcal) are in this dish? 0

Result: Persuasion Without Prior Belief

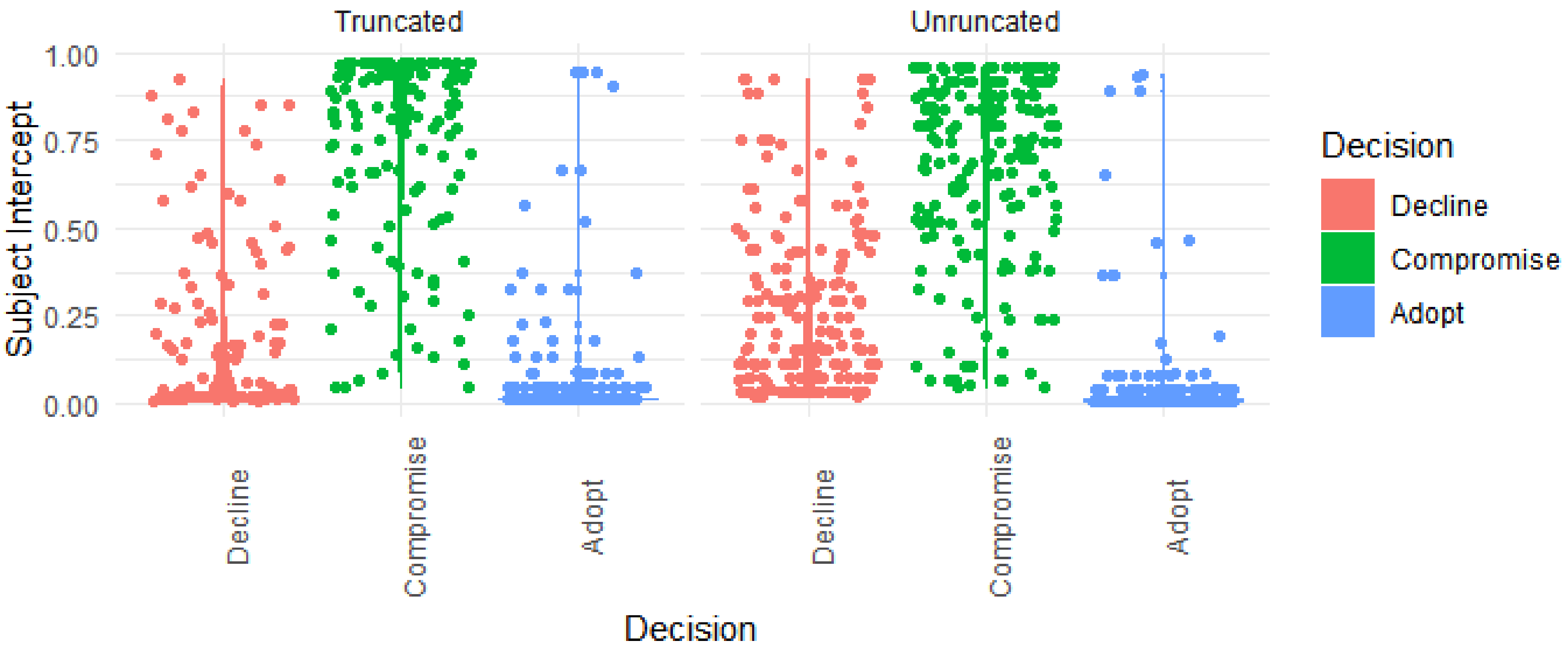








Individual Differences in Decision Strategy



Conclusions

Methodological Insights

- Planned missing data designs can cure or reduce measurement reactivity
- Fully Bayesian imputation reliably recaptures population level parameters in planned missing data designs

Belief Revision

- People do not anchor on their prior when it is not explicitly elicited
- People weigh advice *more* heavily in the absence of measurement reactivity
  - Possible reason for difference from Rader et al. (2015): people believed their personal knowledge was more informative in those tasks than it was for calorie estimation

References

Graham, J. W., Taylor, B. J., Olchowski, A. E., & Cumsille, P. E. (2006). Planned missing data designs in psychological research. *Psychological methods*, *11*(4), 323-343.

Harel, O., Stratton, J., & Aseltine, R. (2015). Designed missingness to better estimate efficacy of behavioral studies-application to suicide prevention trials. *Journal of Medical Statistics and Informatics*, *3*(2).

Himmelstein, M. (2022). Decline, adopt or compromise? A dual hurdle model for advice utilization. *Journal of Mathematical Psychology*, *110*, 102695.

Rader, C. A., Soll, J. B., & Larrick, R. P. (2015). Pushing away from representative advice: Advice taking, anchoring, and adjustment. *Organizational Behavior and Human Decision Processes*, 130, 26–43.

Rhemtulla, M., & Hancock, G. R. (2016). Planned missing data designs in educational psychology research. *Educational Psychologist*, *51*(3-4), 305-316.

Soll, J. B., & Larrick, R. P. (2009). Strategies for revising judgment: How (and how well) people use others' opinions. *Journal of experimental psychology: Learning, memory, and cognition*, *35*(3), 780.