Moral Approach-Avoidance Tendencies as a Measure of Implicit Moral Attitudes and Motivation

Sanskriti Purohit, Natalie Smith, Kristi Van Meter, Paul McKee

MAAT: Do humans exhibit moral approach-avoidance tendencies (and can they serve as proxies for implicit attitudes)?

WELLS FARGO

WELLS FARGO
American
Red Cross

WELLS FARGO

R-MAAT: To what extent does race interact with moral approach-avoidance tendencies?

WELLS FARGO

WELLS FARGO

What we know so far: It is unclear

What have we done so far: MAAT Study 1

MAAT

- Stimuli are either: 3-word **Phrases** or **Images**.
- Stimuli are two of three **Moral Valence** categories:
  - Good and Bad , Good and Neutral, or Neutral and Bad
- Stimuli **Background Color** are either:
  - Yellow or Blue
- Stimuli are **Evaluated** by either:
  - **Implicit** (responding to Background Color) or **Explicit** (responding to Moral Valence)

Moral Valence Examples

Morally Good

Example: Saving a life

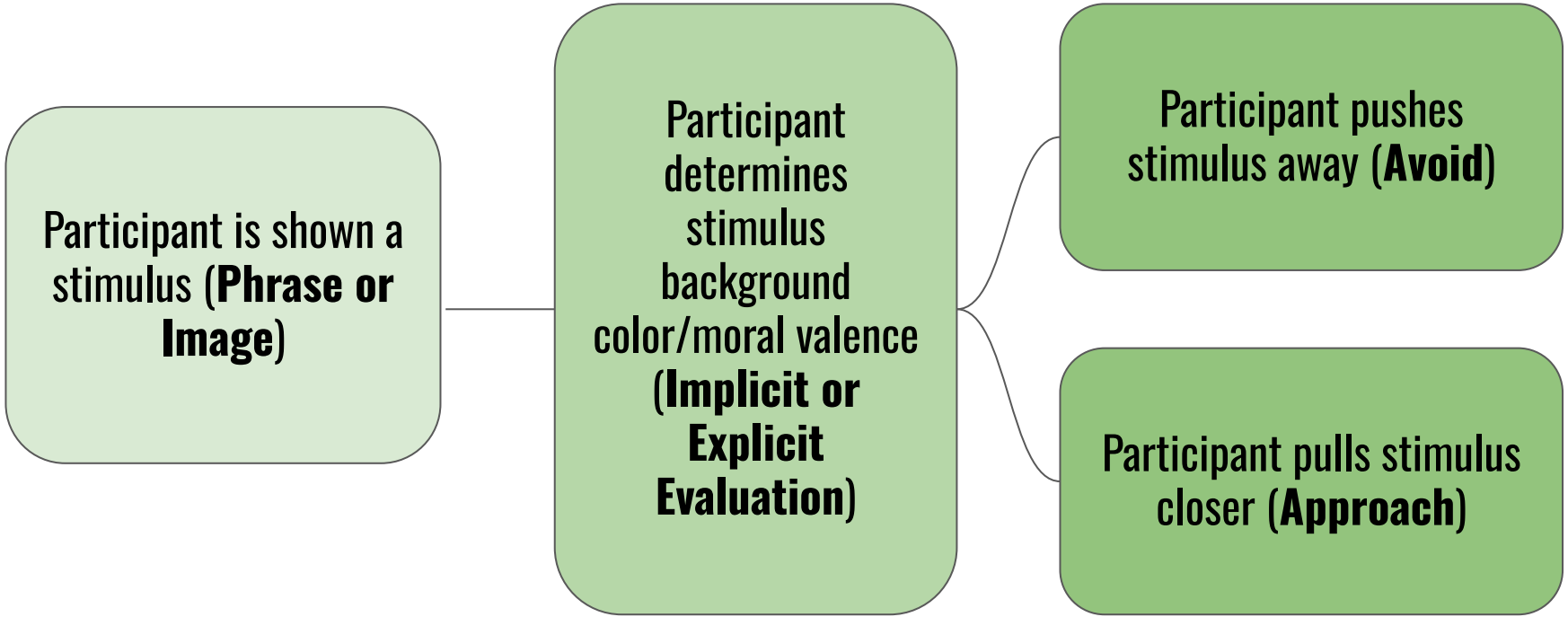
Morally Neutral

Example: Eating a pizza, Stubbing a toe

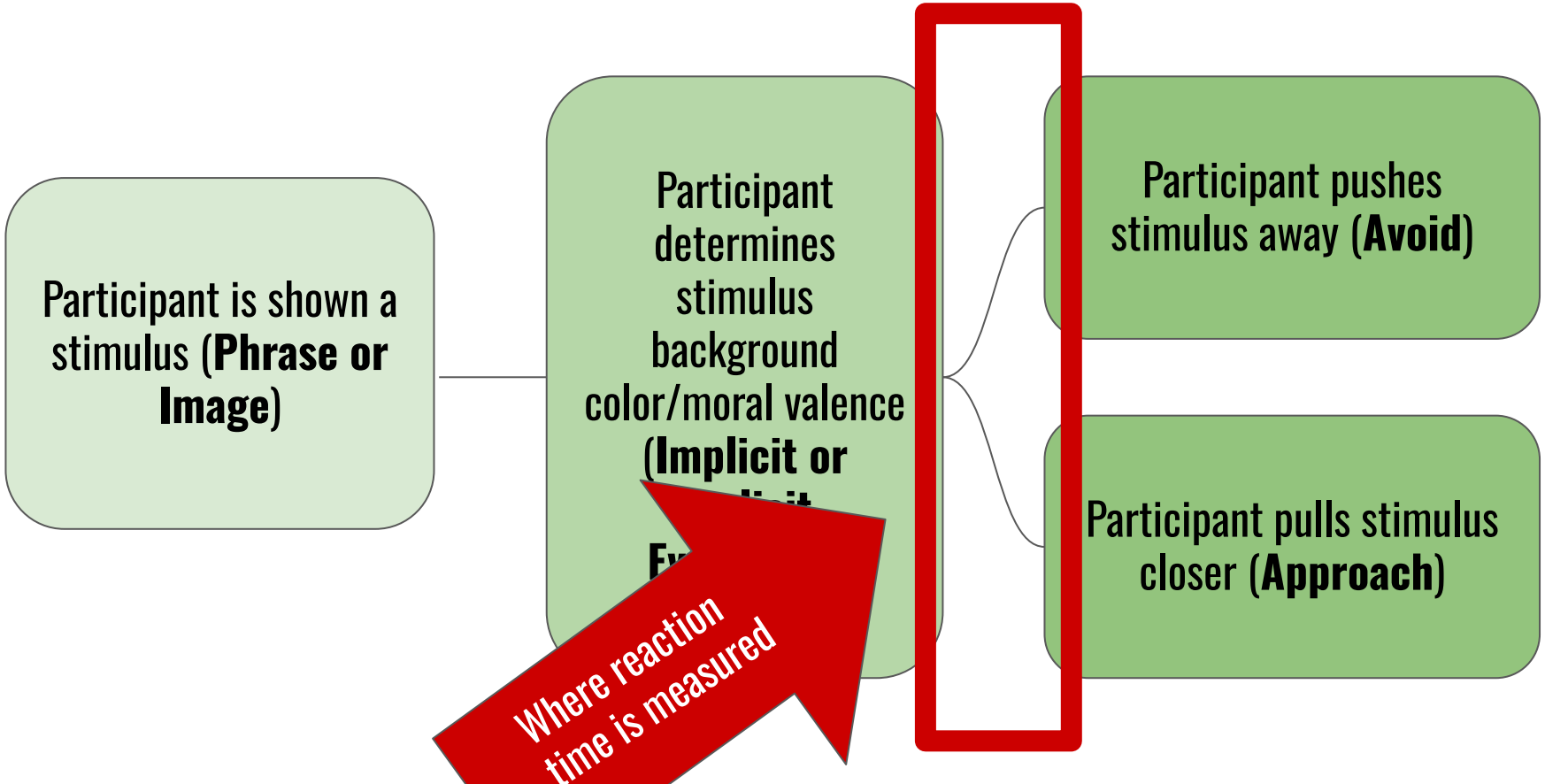
Morally Bad

Example: Robbing a store

The Approach Avoidance Task (AAT)



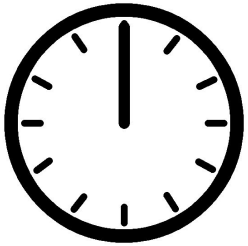
The Approach Avoidance Task (AAT)

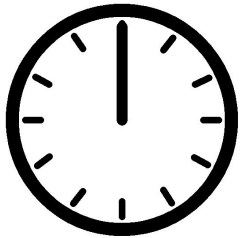


Appendix

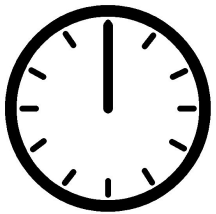
Linear Mixed Effects Models

Linear Mixed Effects Models

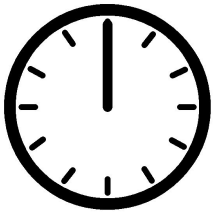




Linear Mixed Effects Models



Linear Mixed Effects Models

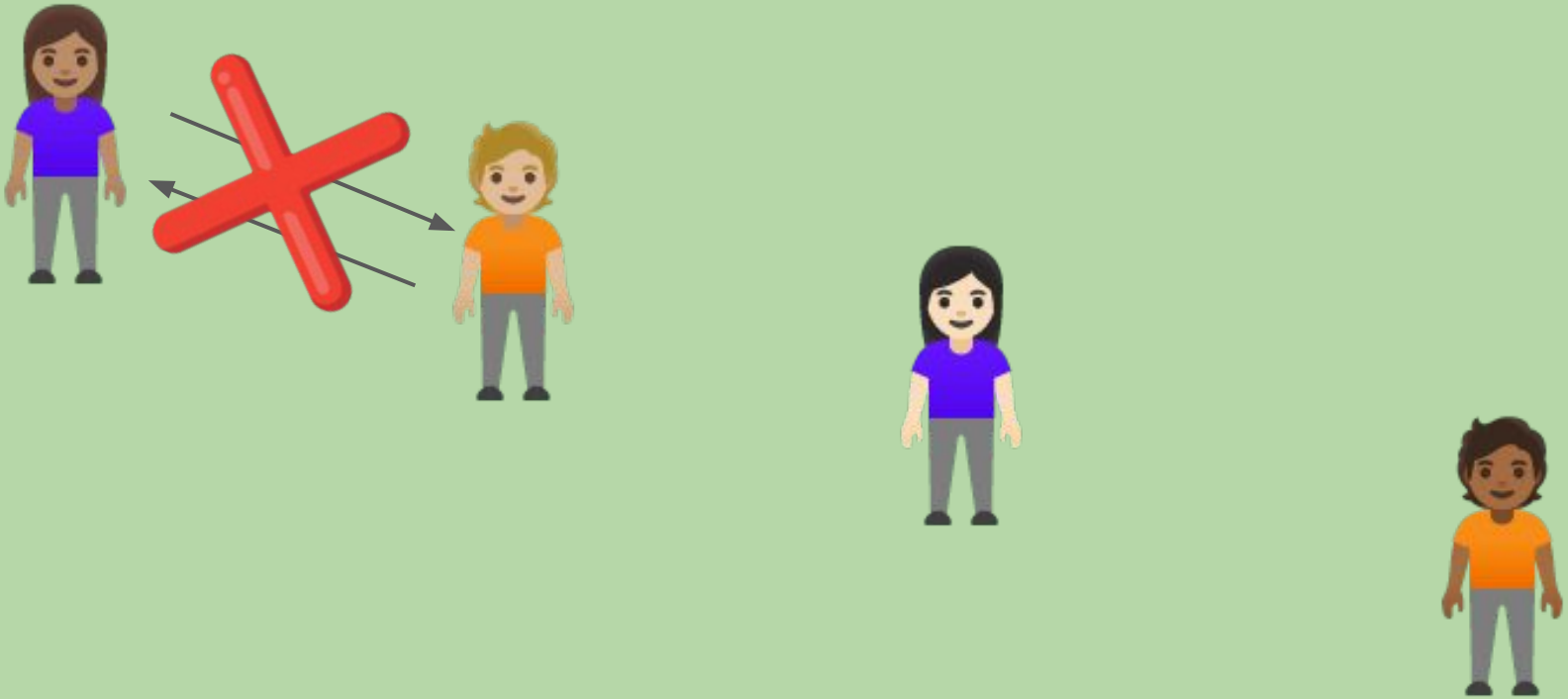


LINEAR MODELS

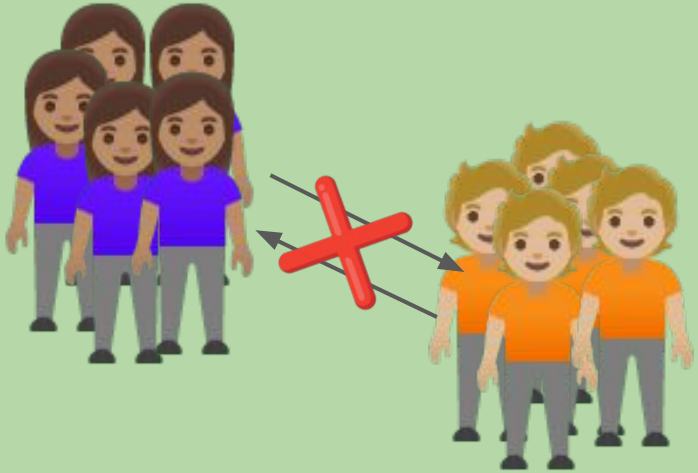
$$\text{Outcome} = E(\text{Outcome}) + AX + BY + \text{Error}$$

Linear Mixed Effects Models

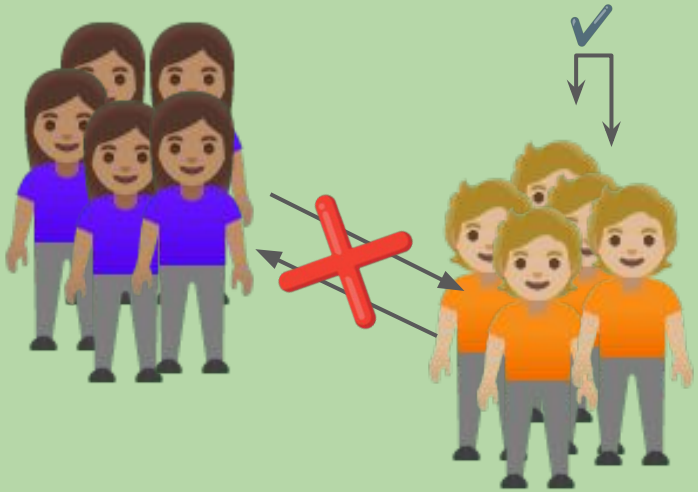
Linear Mixed Effects Models



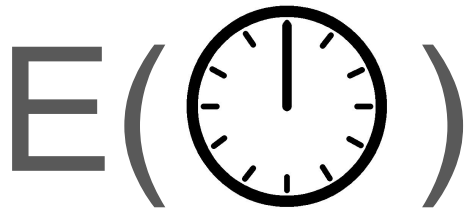
Linear Mixed Effects Models



Linear Mixed Effects Models



Linear Mixed Effects Models



Linear Mixed Effects Models

E(🕐) +

Linear Mixed Effects Models

LINEAR MIXED EFFECTS MODELS

$$Outcome_i = E(Outcome)_i + AX + BY + Error$$

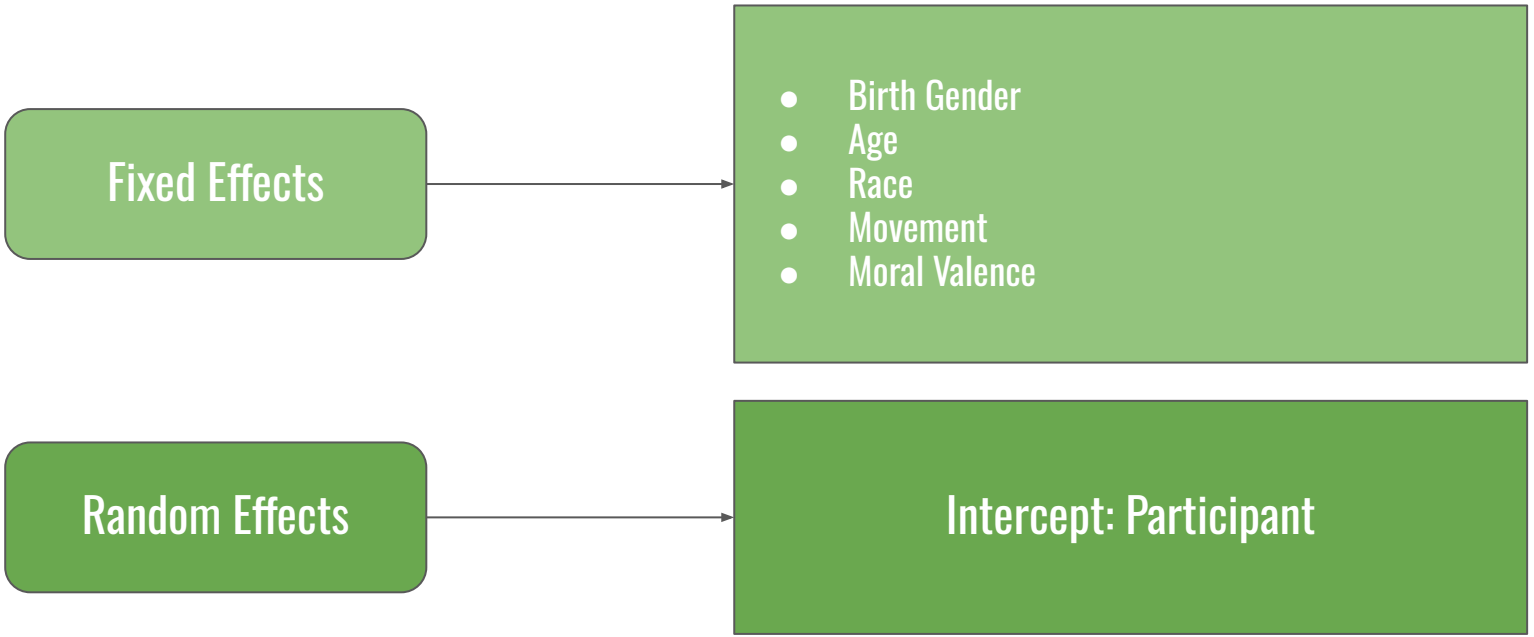
LINEAR MODELS

$$Outcome_i = E(Outcome)_i + AX + BY + Error$$

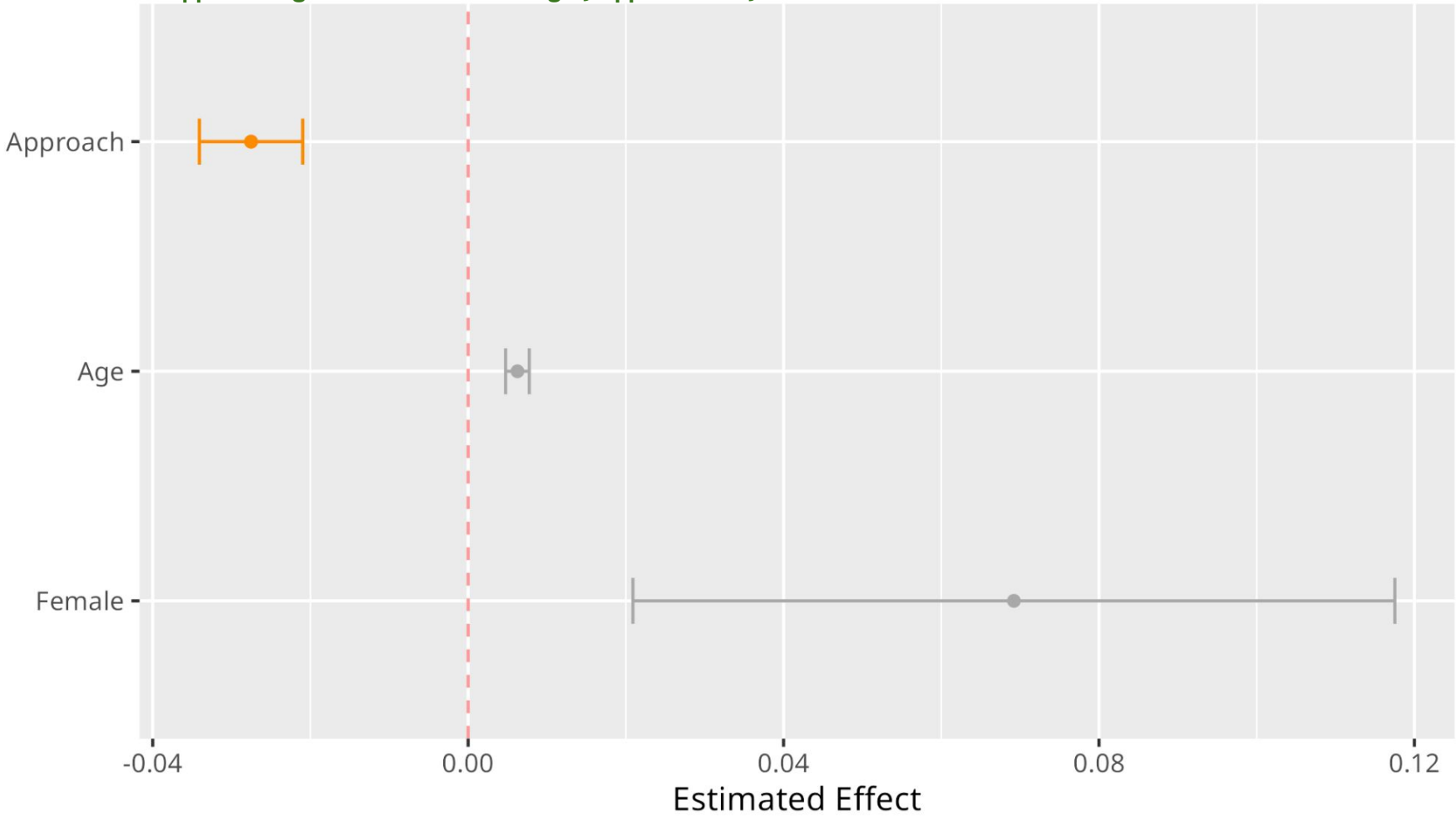
$$E(Outcome)_i = E(Outcome) + \text{ }_i$$

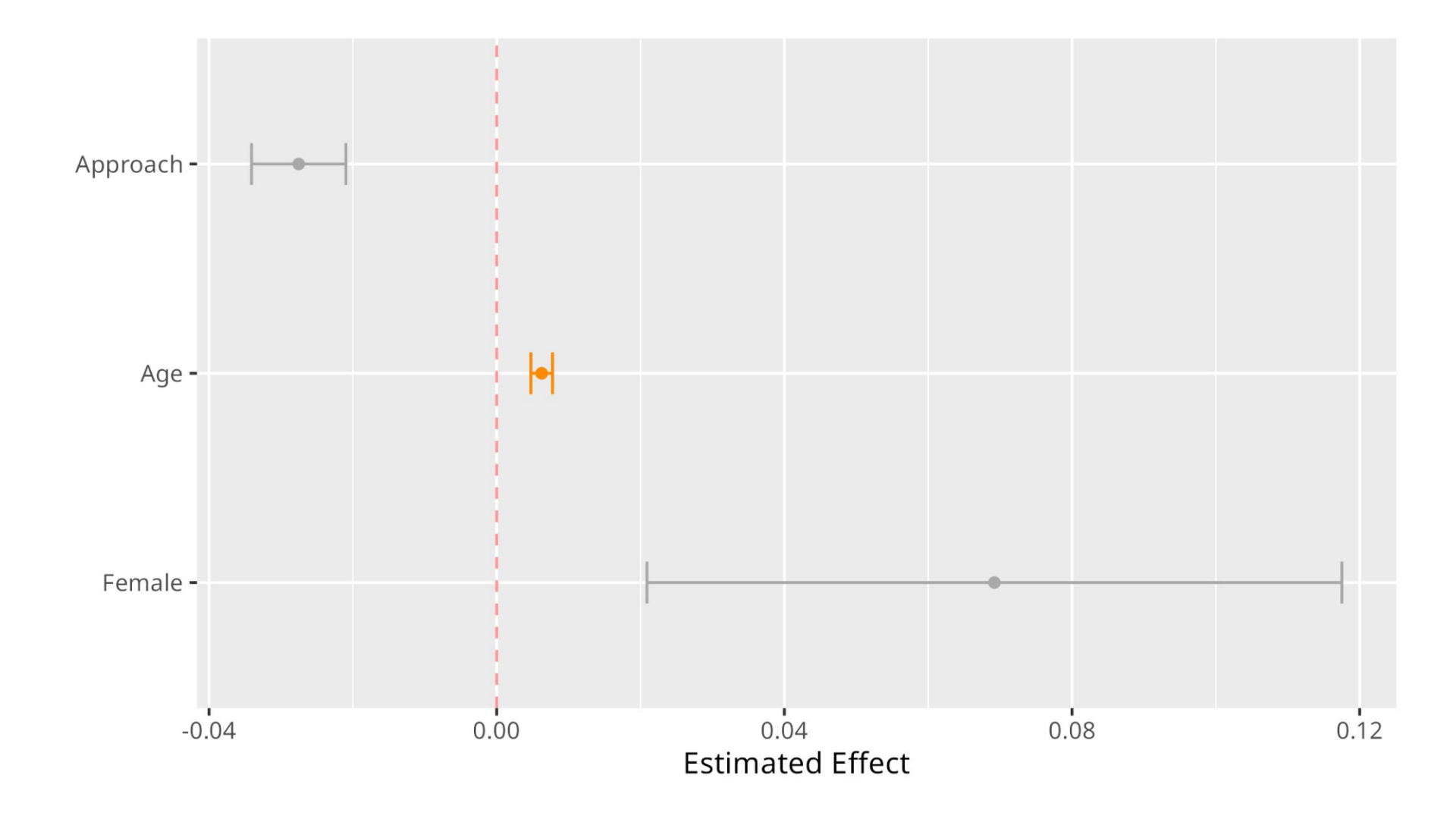
MODEL RESULTS

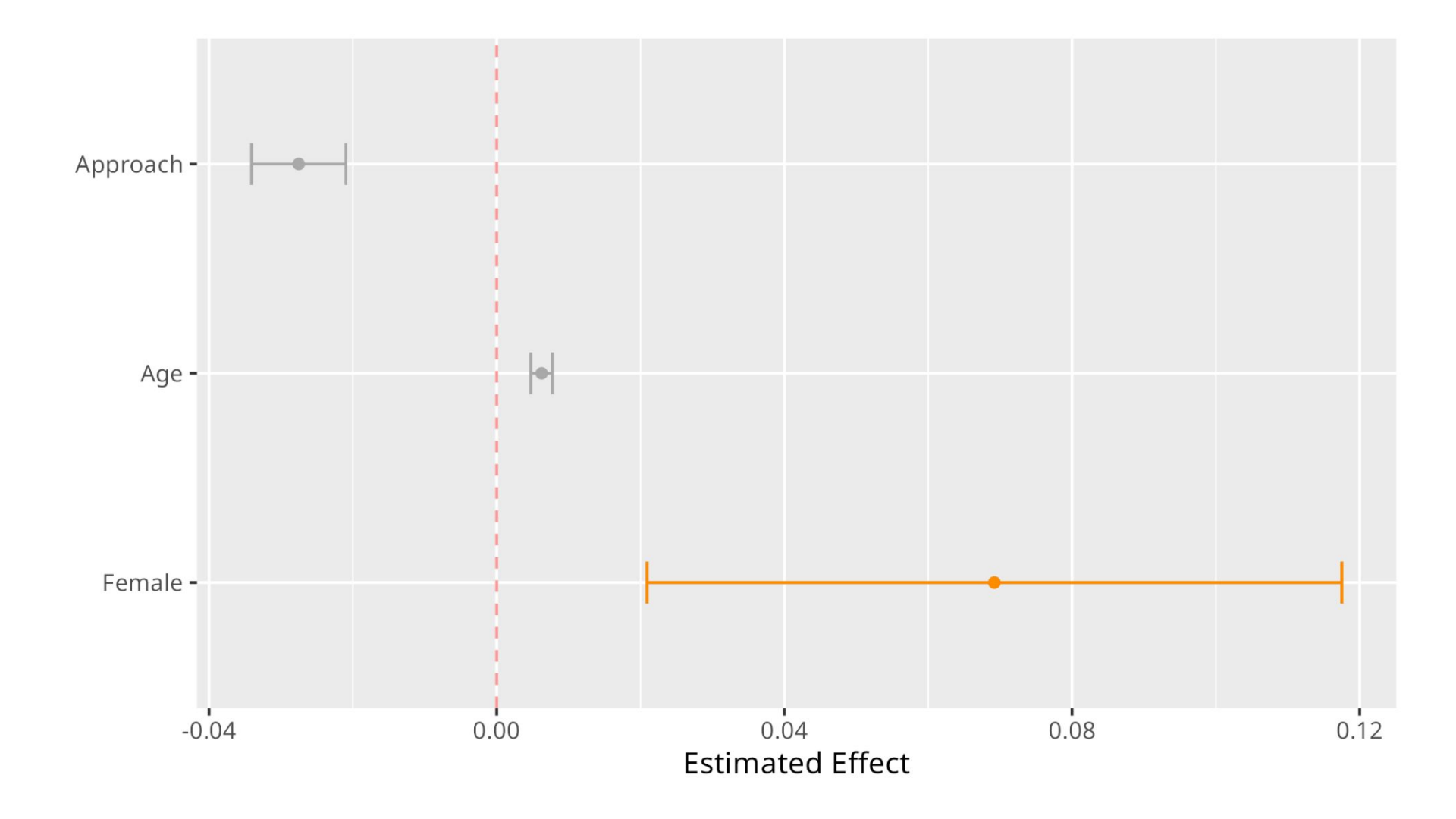
Linear Mixed Effects Models: Implicit (Phrases)



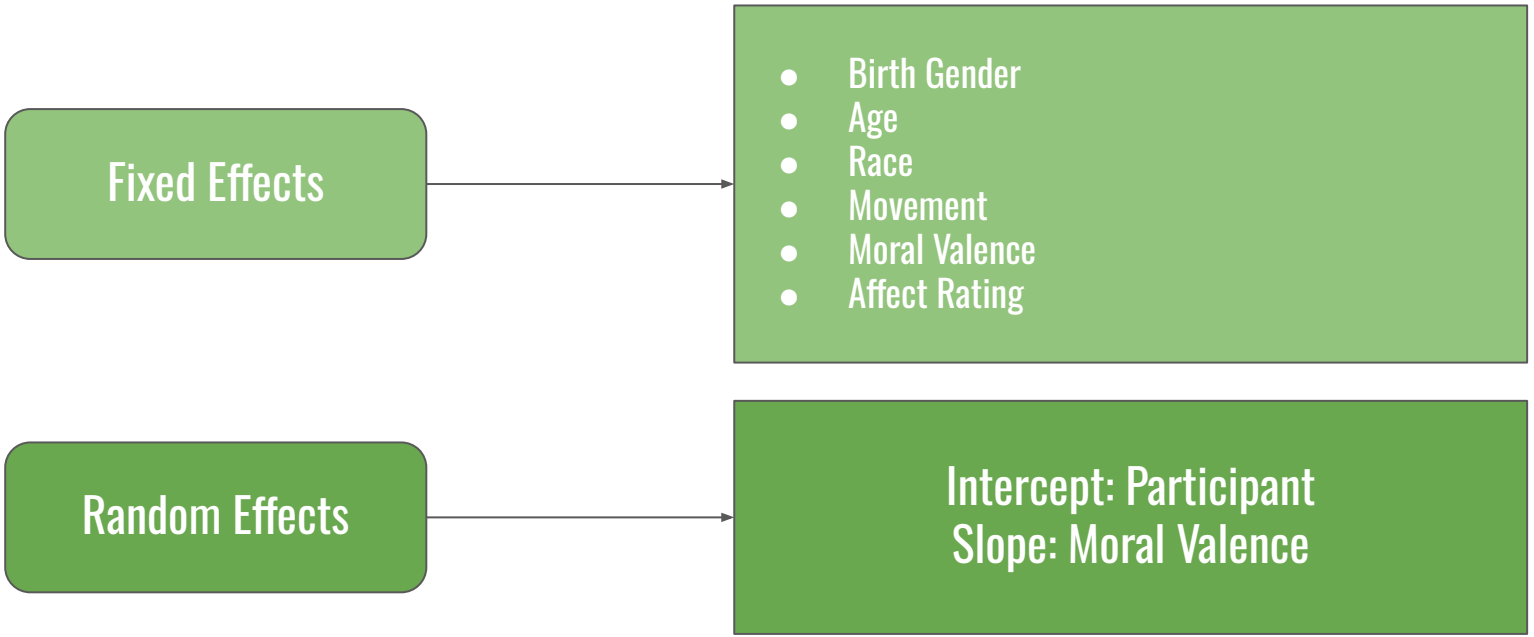
Approaching is faster than avoiding by approximately 3%

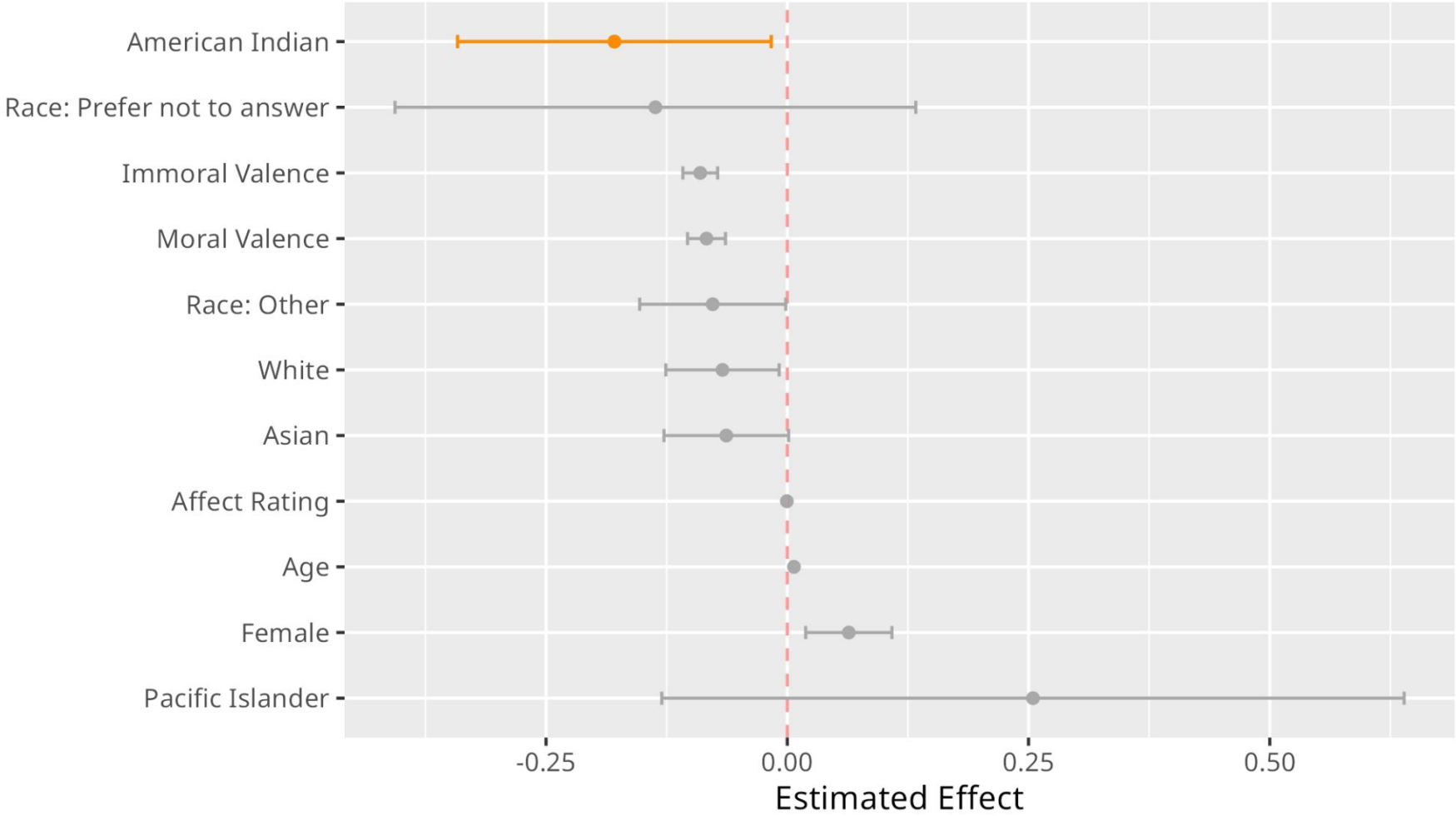



Approach
Age
Female
-0.04
0.00
0.04
0.08
0.12
Estimated Effect

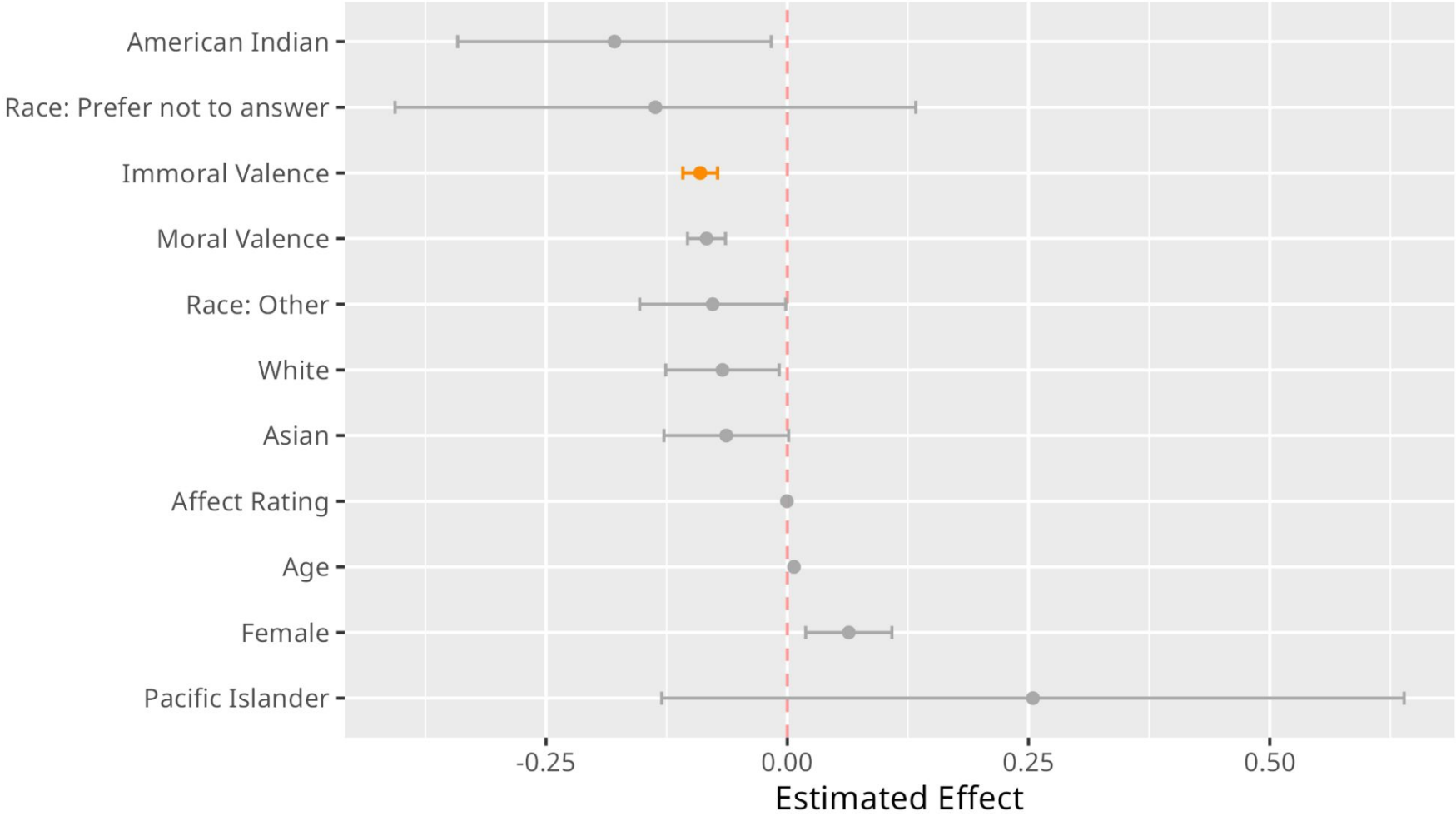

Approach
Age
Female
-0.04
0.00
0.04
0.08
0.12
Estimated Effect

Linear Mixed Effects Models: Explicit (Phrases)

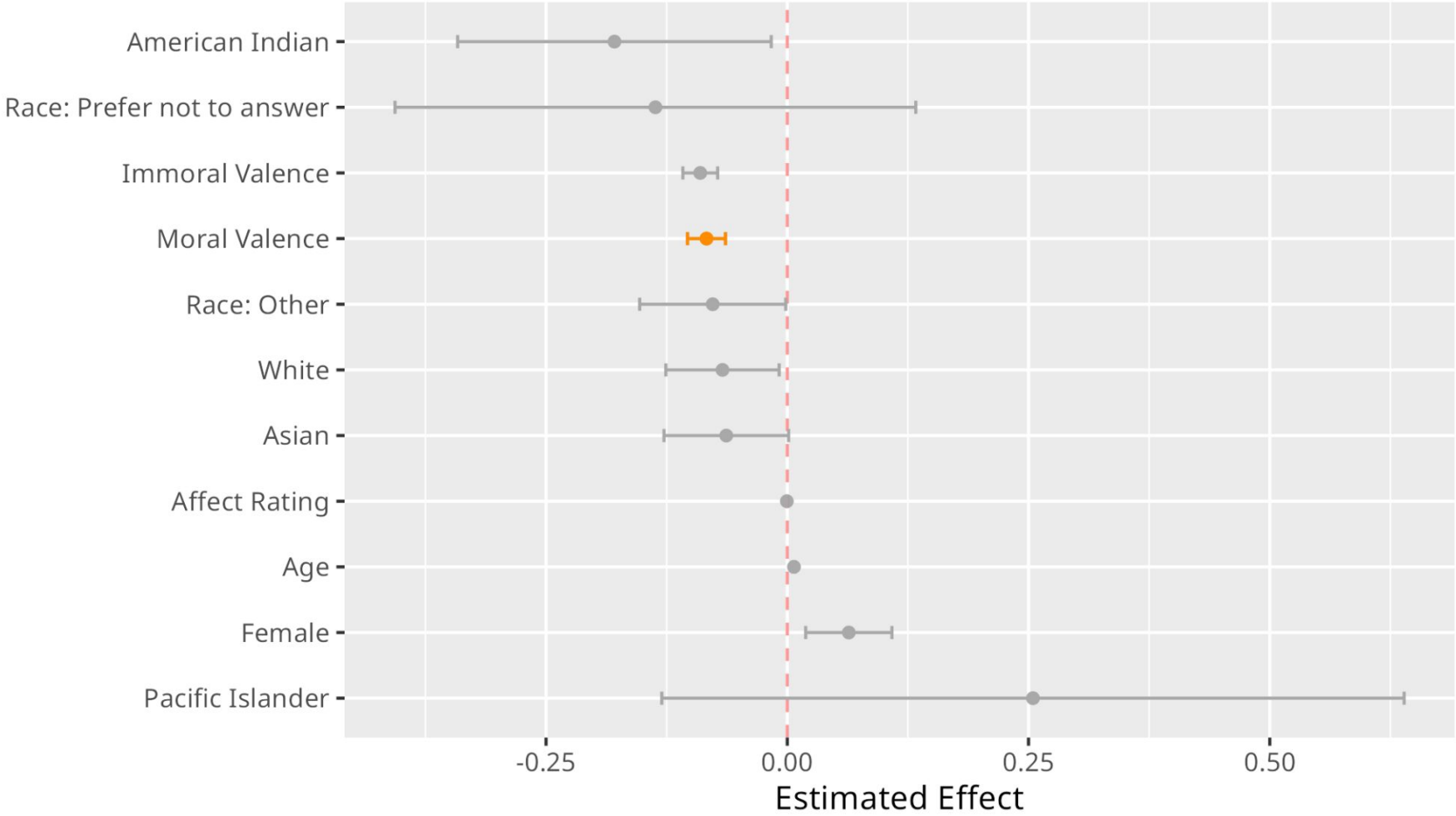



American Indian
Race: Prefer not to answer
Immoral Valence
Moral Valence
Race: Other
White
Asian
Affect Rating
Age
Female
Pacific Islander
-0.25
0.00
0.25
0.50
Estimated Effect

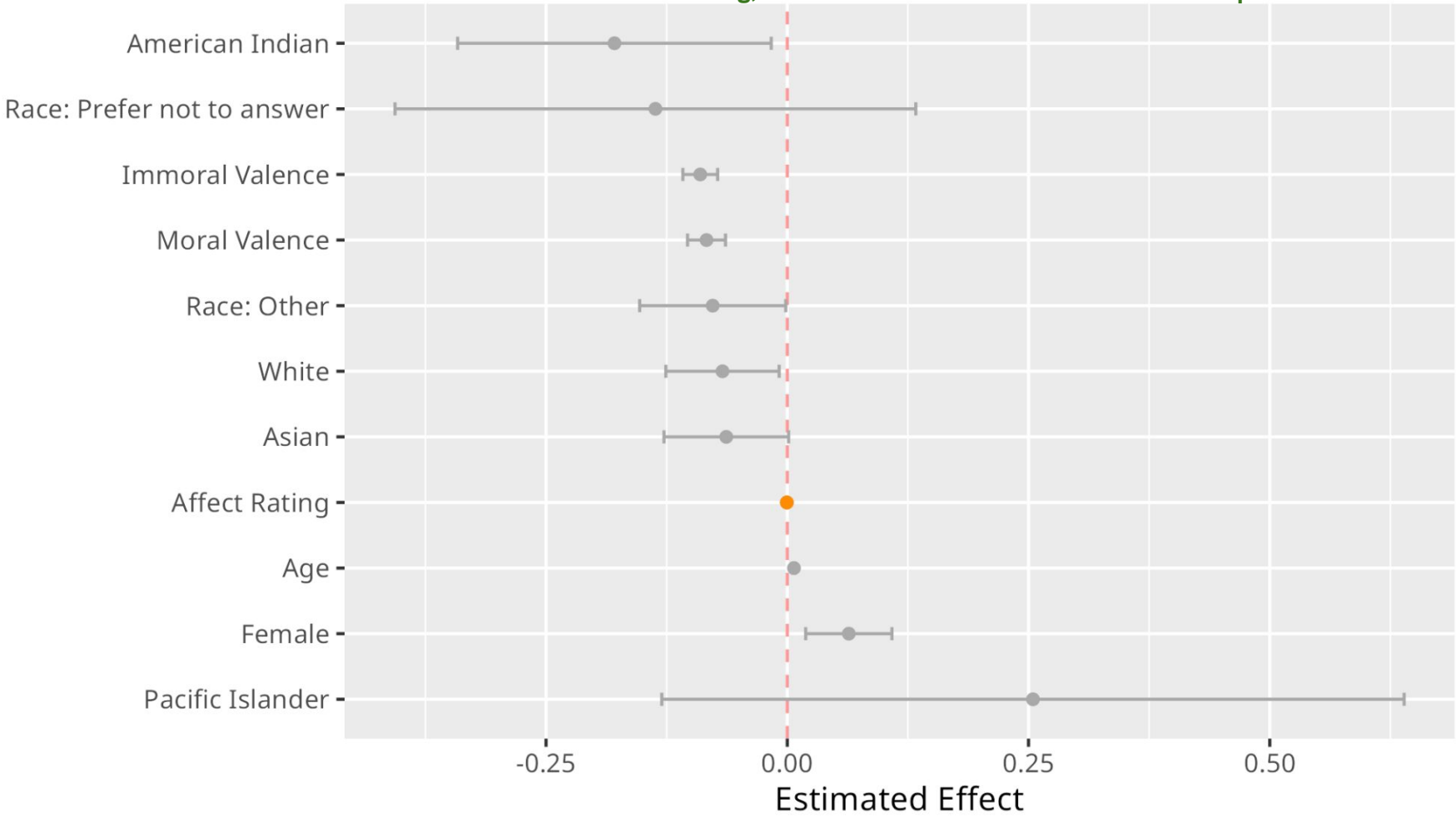
On average, compared to Neutral-rated phrases, responses to Immoral-rated phrases are faster by 8.6%.

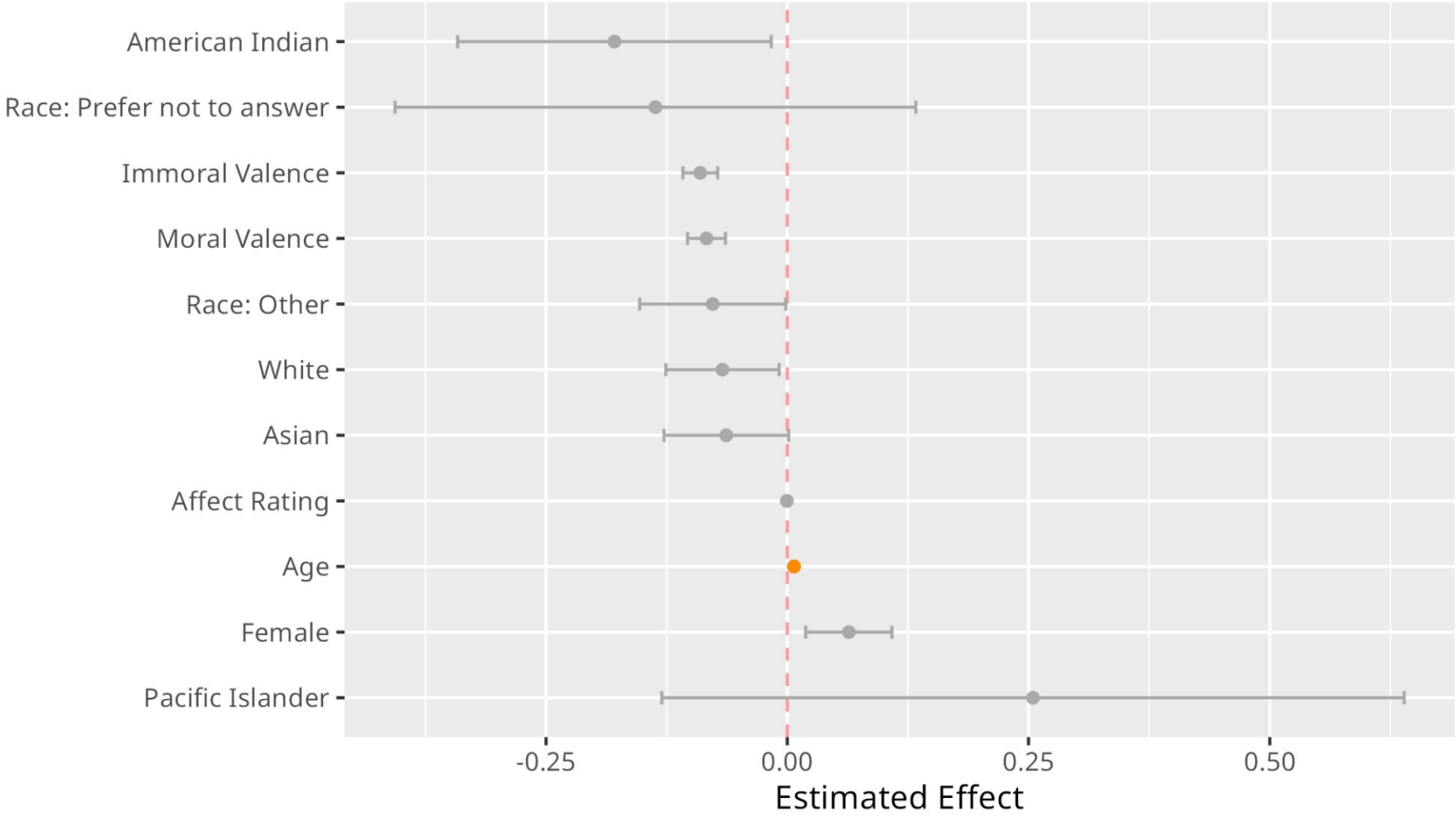


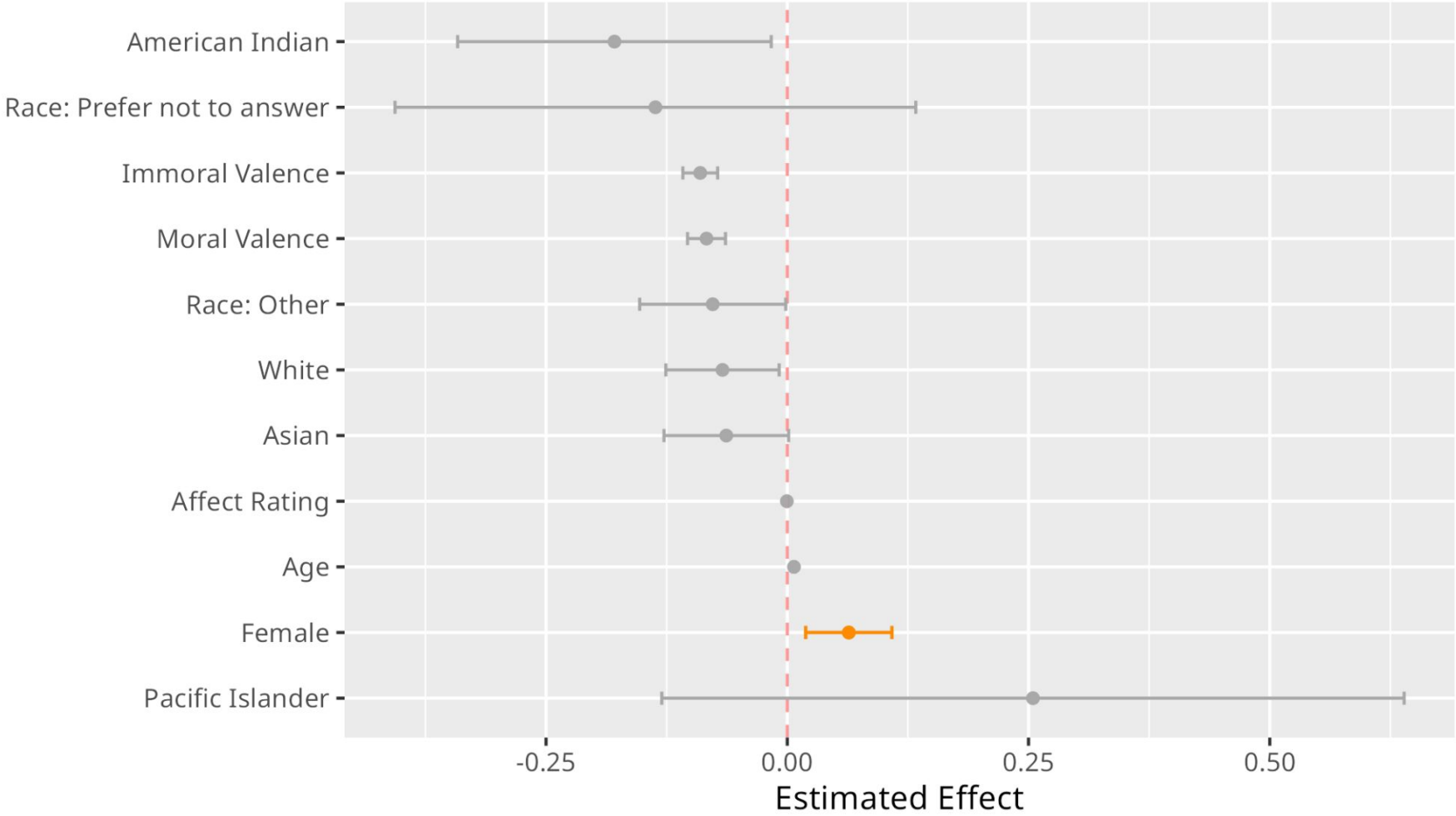
On average, compared to Neutral-rated phrases, responses to Moral-rated phrases are faster by 7.6%.



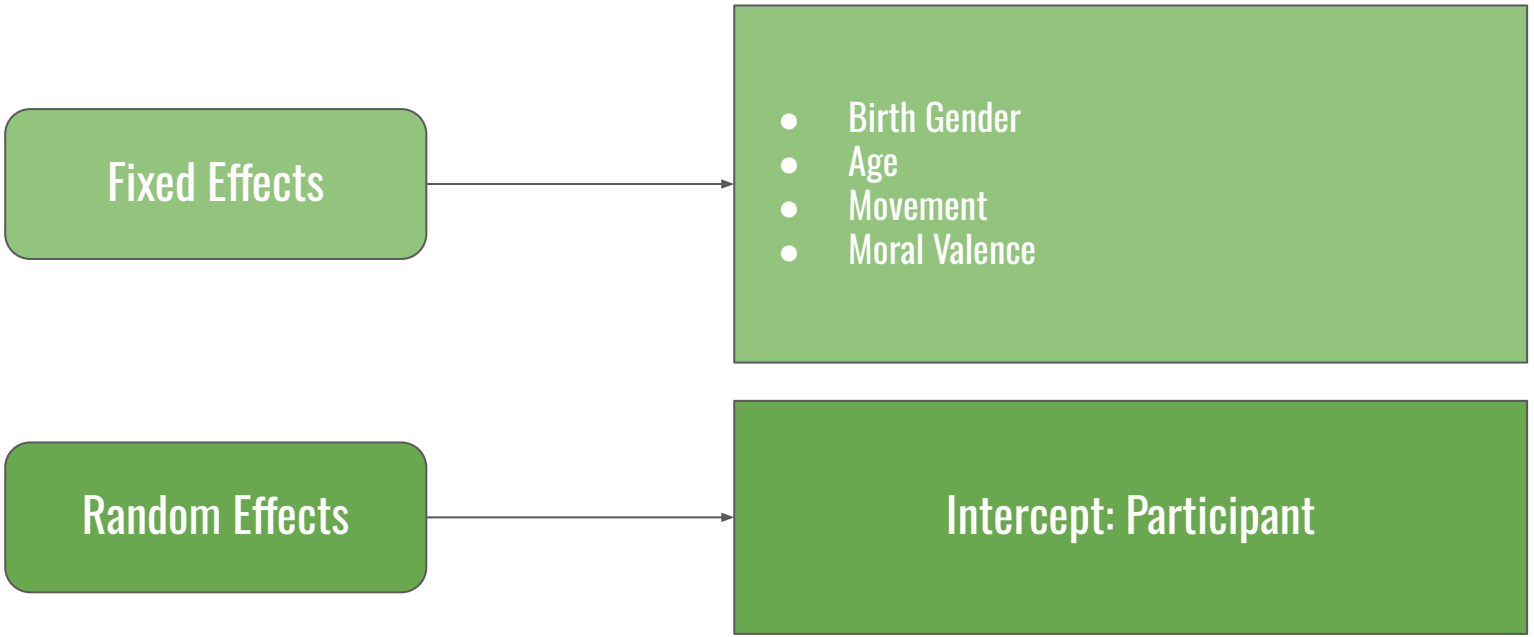
With 1 unit increase in Affect Rating, there is a 1 millisecond increase in the response time.



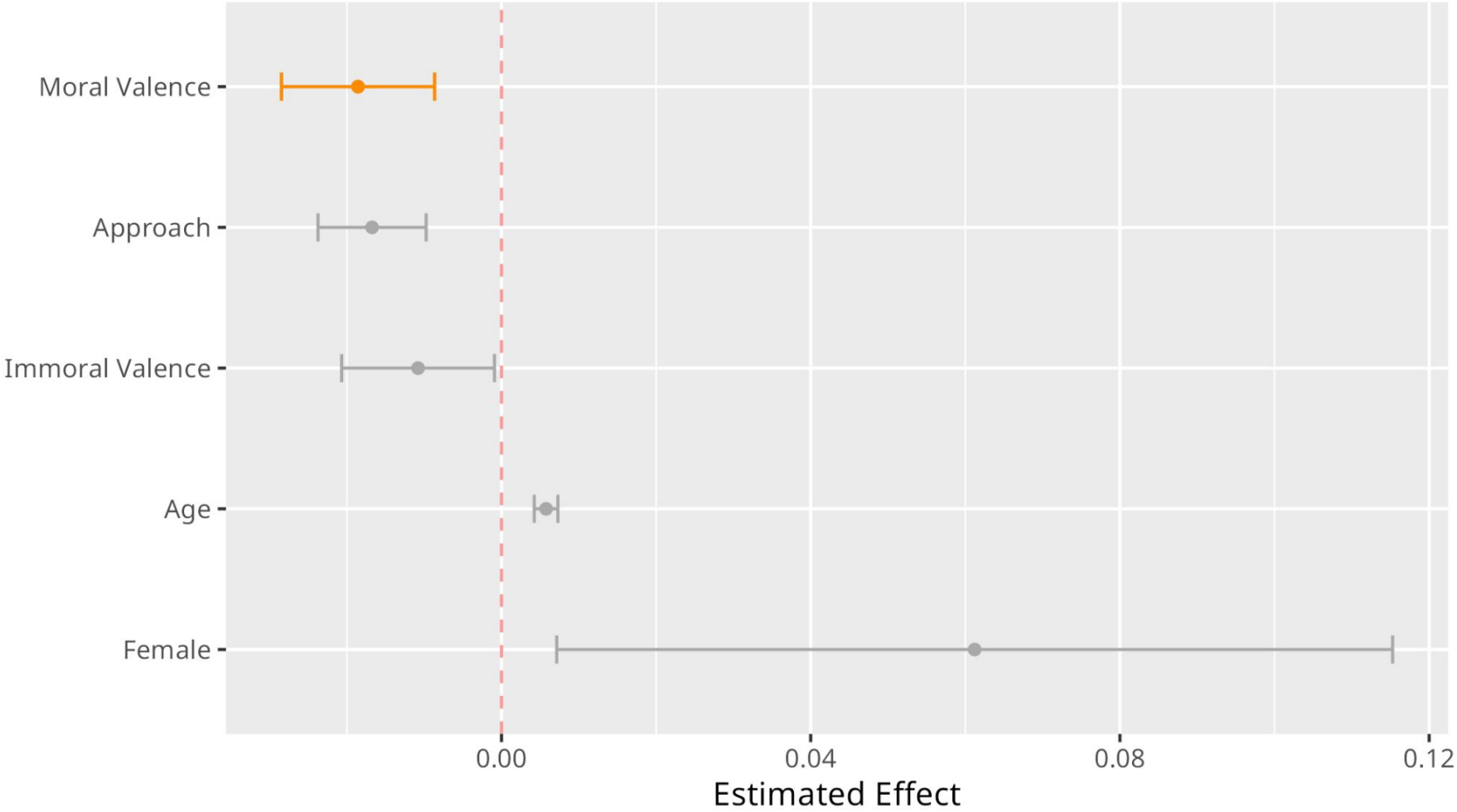

American Indian
Race: Prefer not to answer
Immoral Valence
Moral Valence
Race: Other
White
Asian
Affect Rating
Age
Female
Pacific Islander
-0.25
0.00
0.25
0.50
Estimated Effect


American Indian
Race: Prefer not to answer
Immoral Valence
Moral Valence
Race: Other
White
Asian
Affect Rating
Age
Female
Pacific Islander
-0.25
0.00
0.25
0.50
Estimated Effect

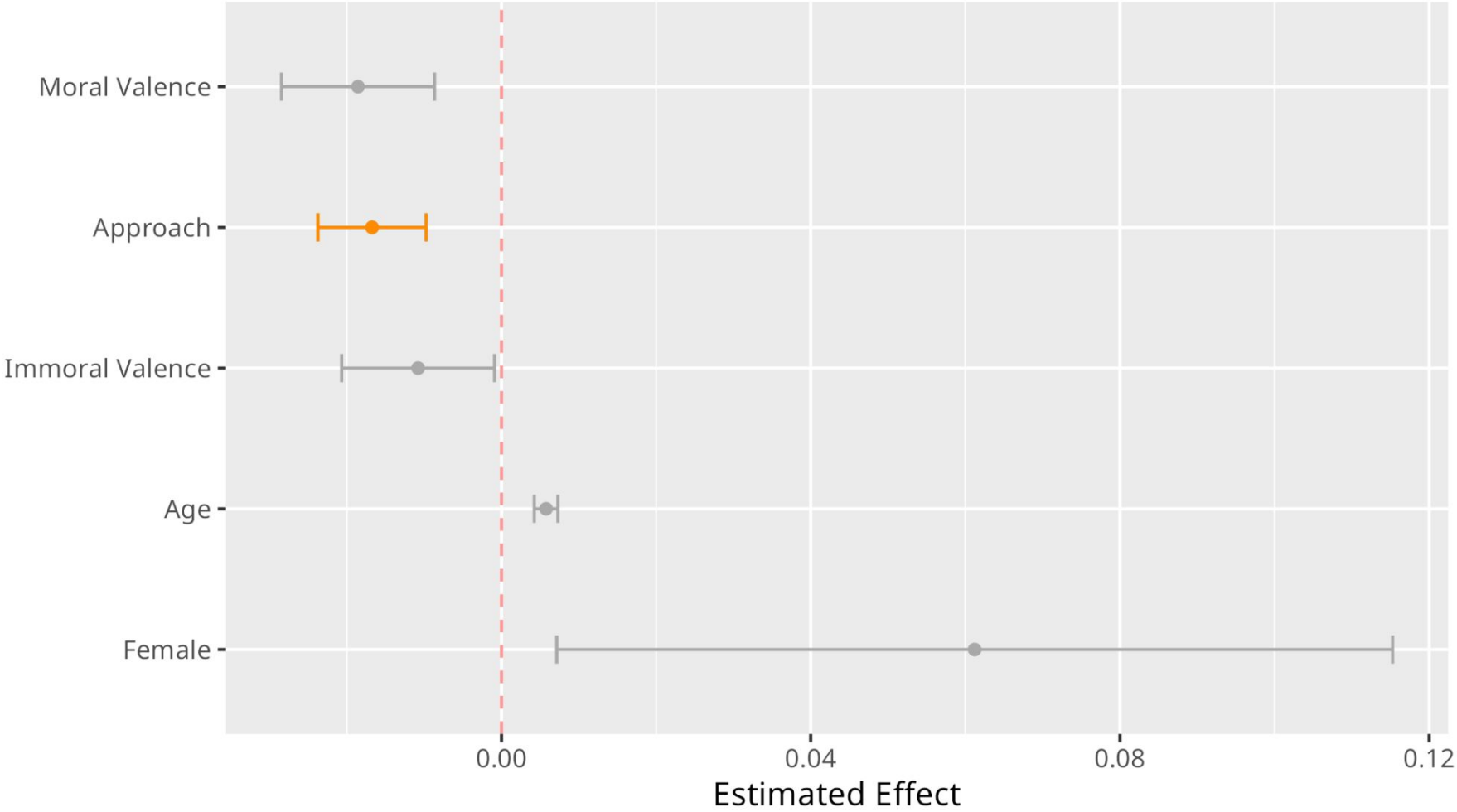
Linear Mixed Effects Models: Implicit (Images)



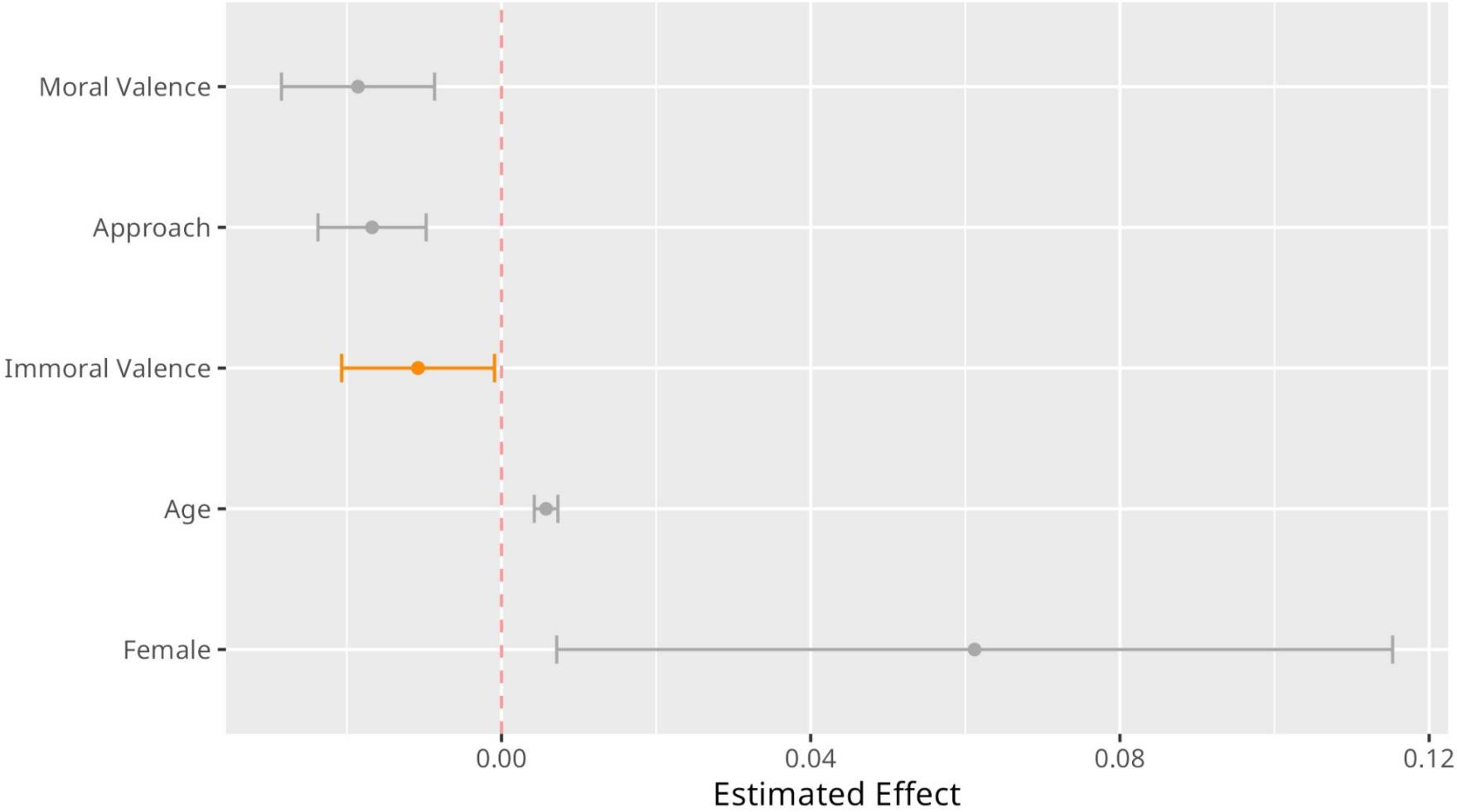
On average, compared to Neutral-rated phrases, responses to Moral-rated phrases are faster by 1.9%.

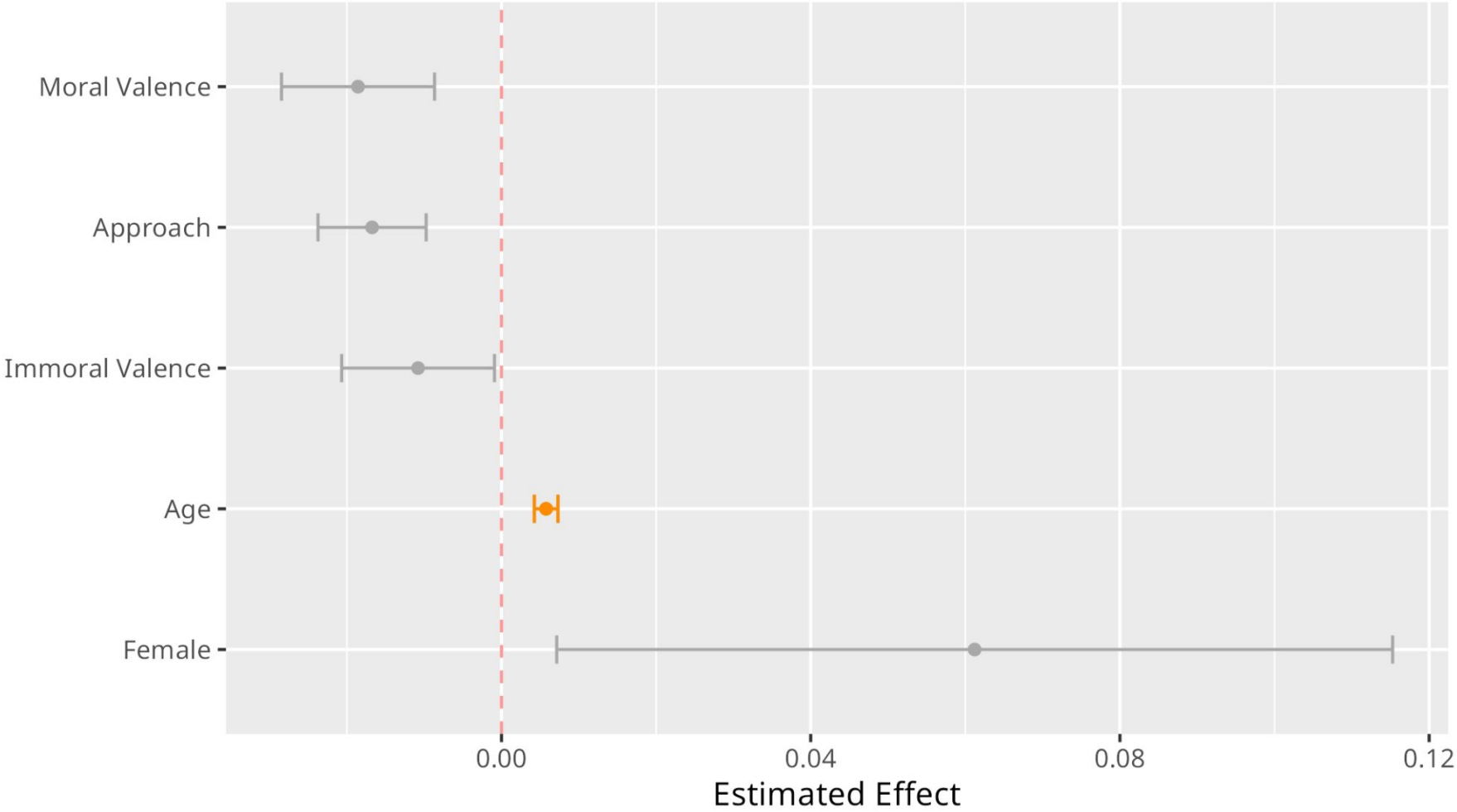


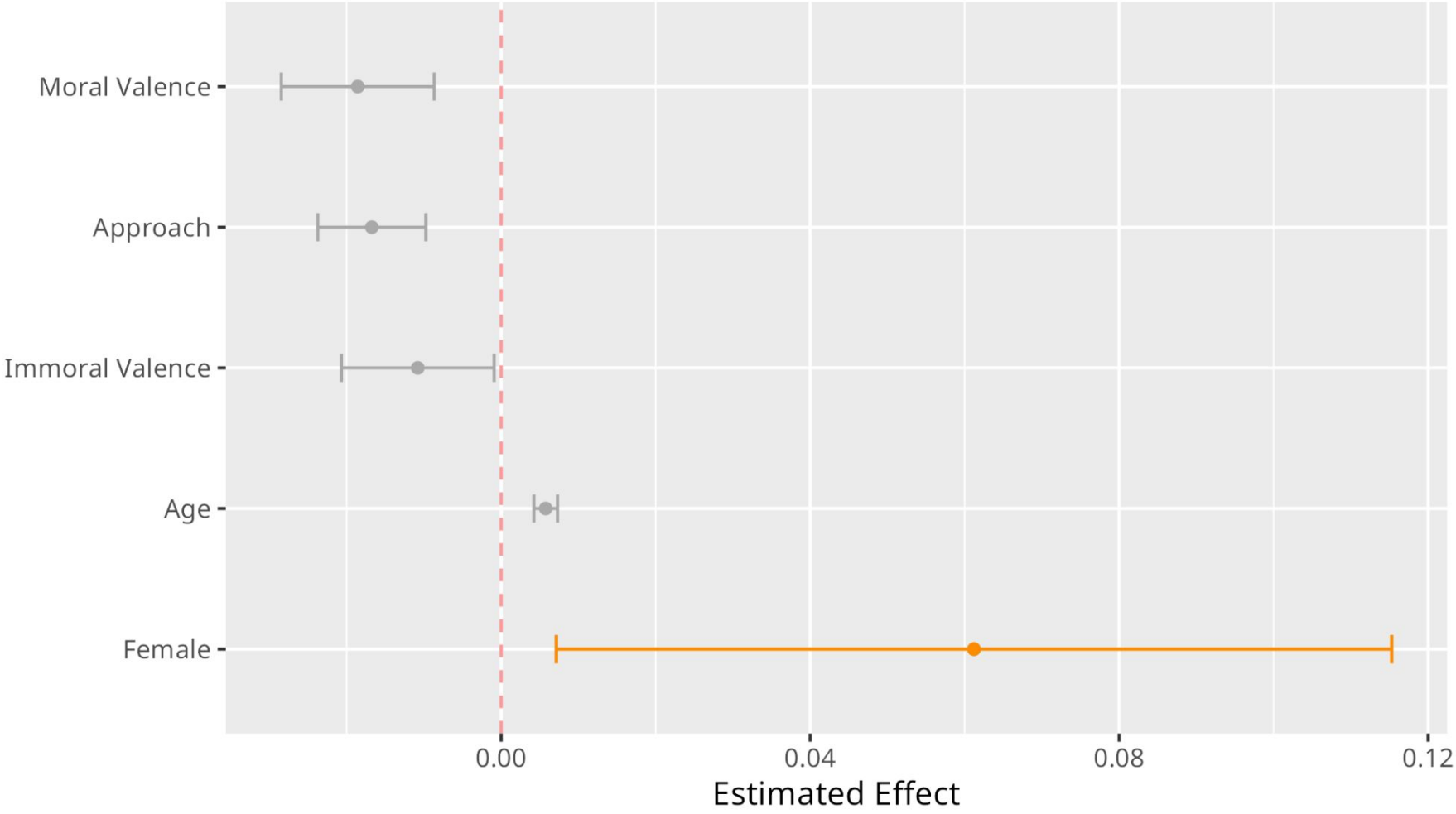
Approaching is faster than avoiding by approximately 1.6% (a decrease compared to phrases)



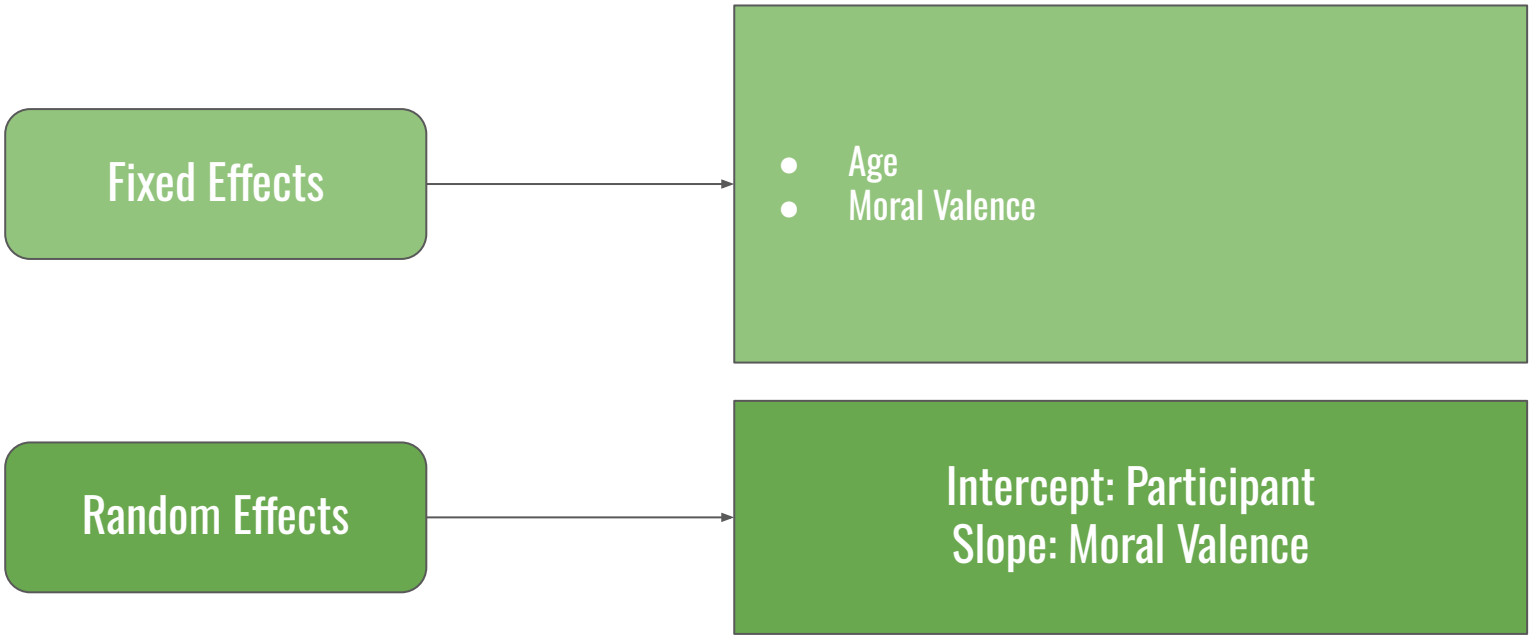
On average, compared to Neutral-rated phrases, responses to Immoral-rated phrases are faster by 1%.



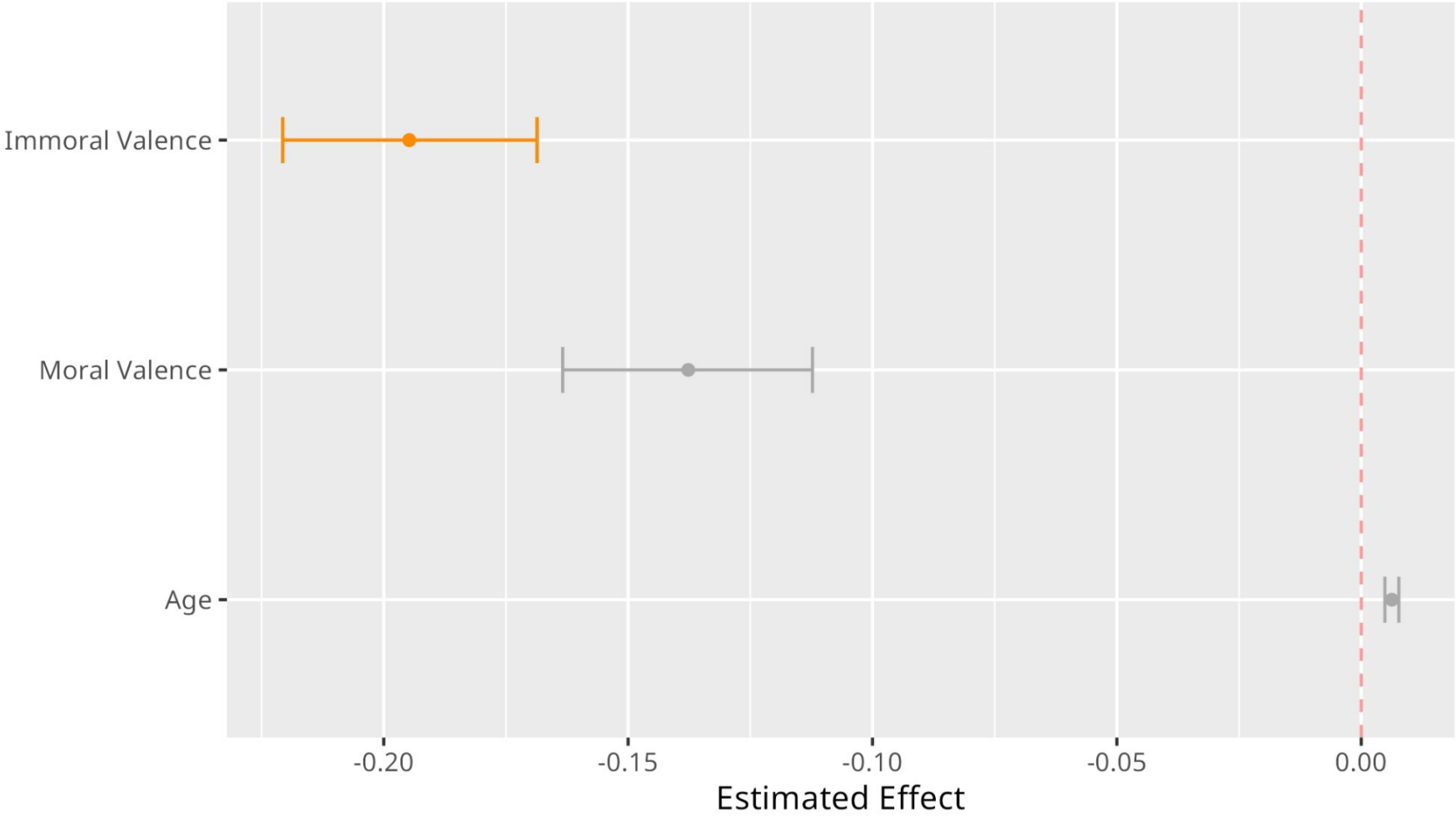

Moral Valence
Approach
Immoral Valence
Age
Female
0.00
0.04
0.08
0.12
Estimated Effect


Moral Valence
Approach
Immoral Valence
Age
Female
0.00
0.04
0.08
0.12
Estimated Effect

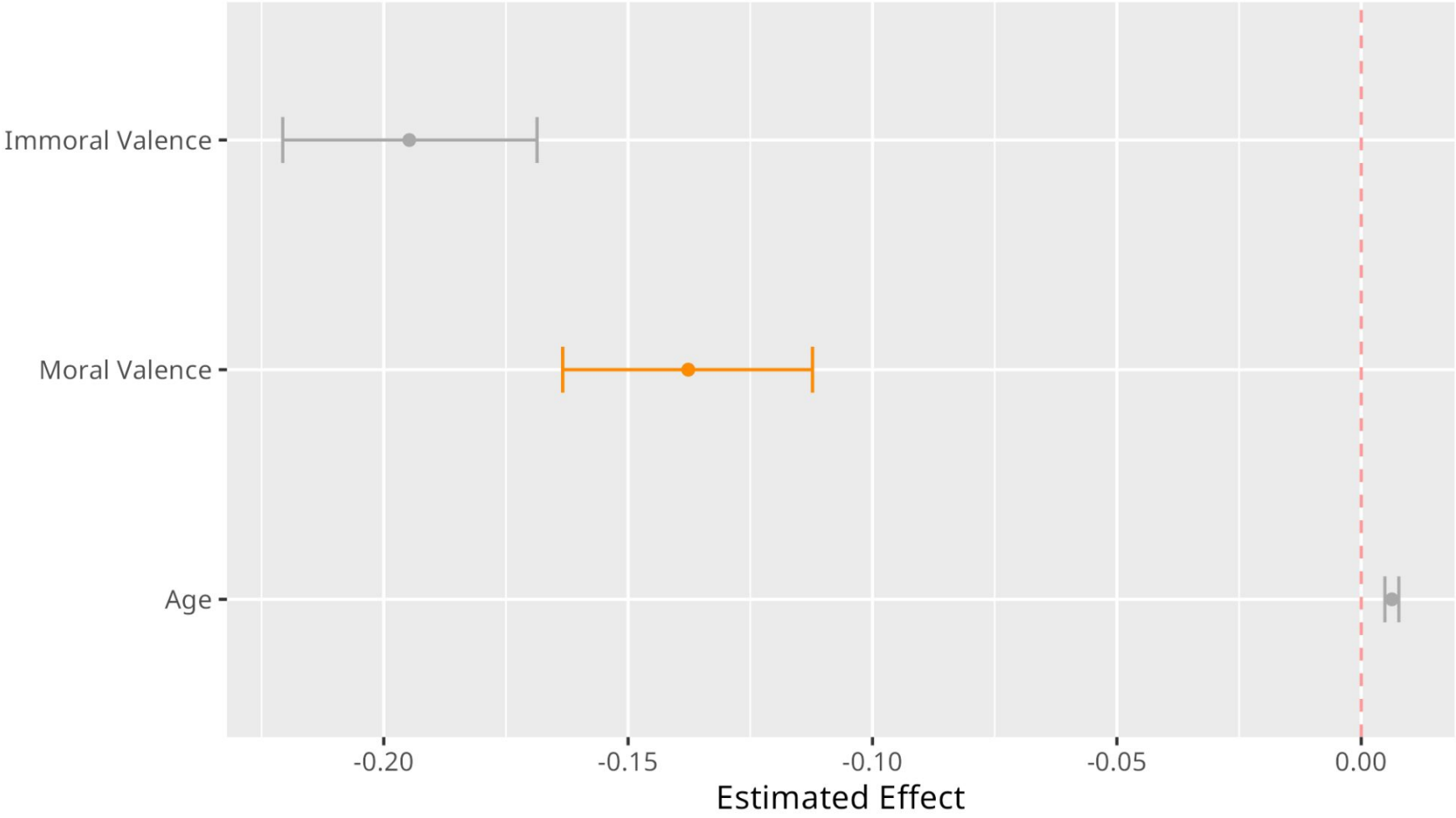
Linear Mixed Effects Models: Explicit (Images)

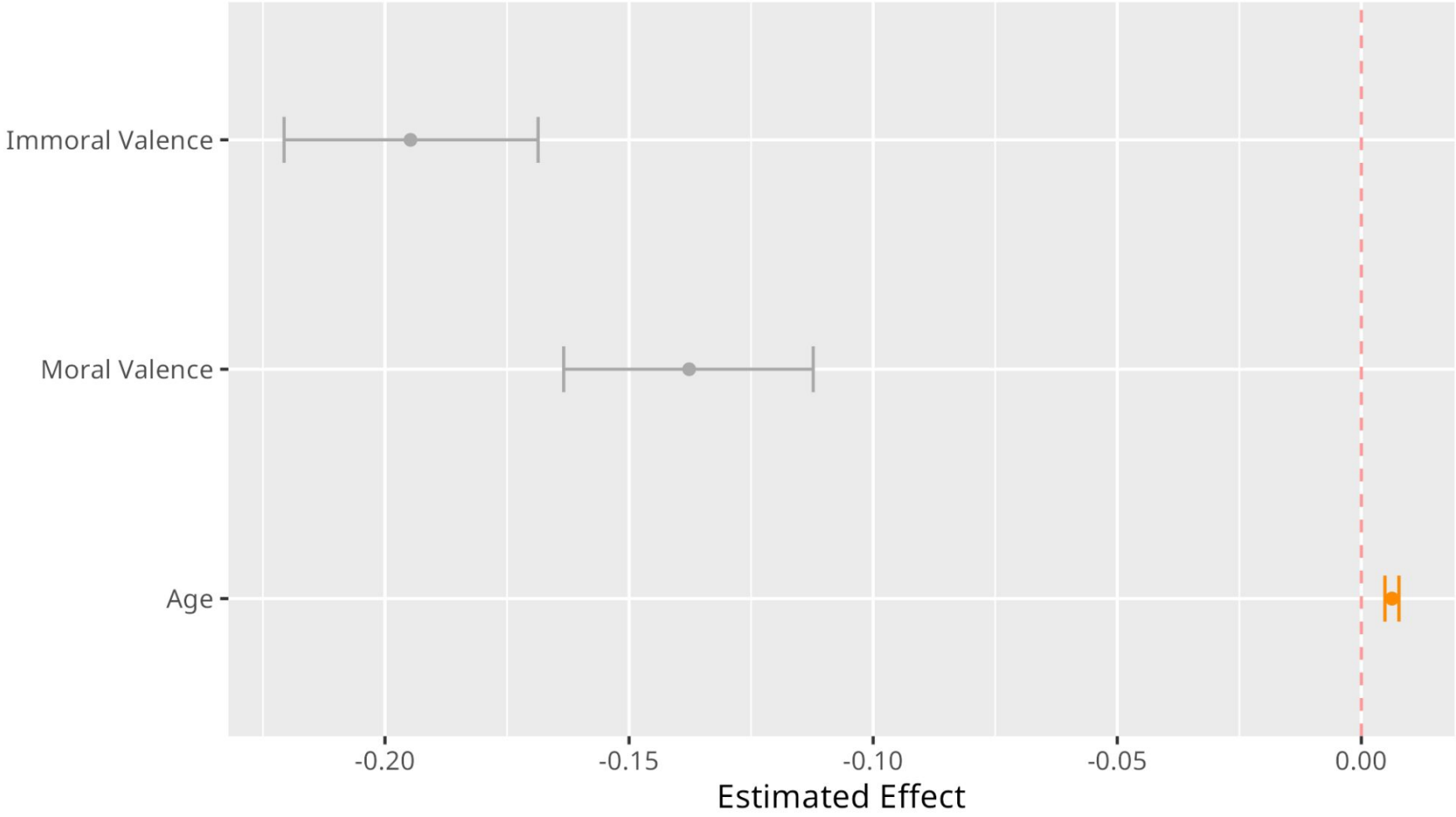


On average, compared to Neutral-rated phrases, responses to Immoral-rated phrases are faster by 17.3%.



On average, compared to Neutral-rated phrases, responses to Moral-rated phrases are faster by 12.9%.

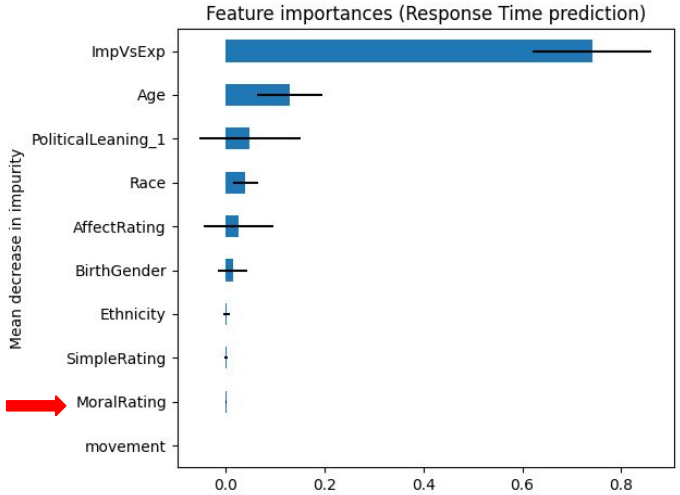



Immoral Valence
Moral Valence
Age
-0.20
-0.15
-0.10
-0.05
0.00
Estimated Effect

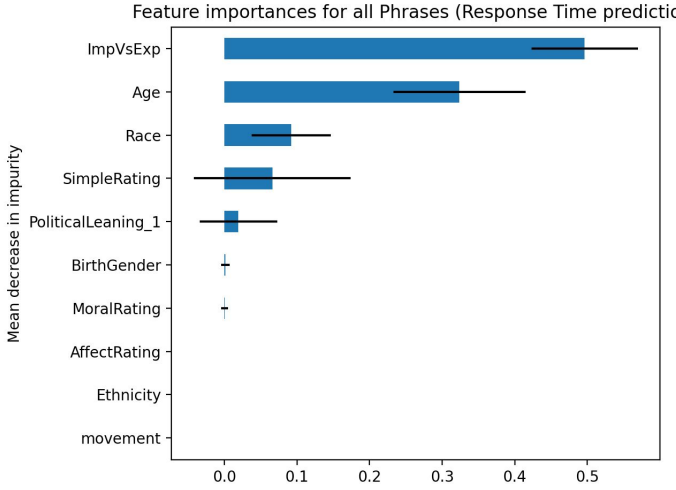
There might be a relationship, but is it predictive?:
Supervised ML Prediction of Response Time

Performance is Poor (particularly for images) and Moral Rating Does Little to Nothing in Predicting Response Time

	R2	MAE
Ridge	0.069	383.15
RF	0.072	340.67
SVR	0.143	292.59
XGBoost	-0.034	335.63

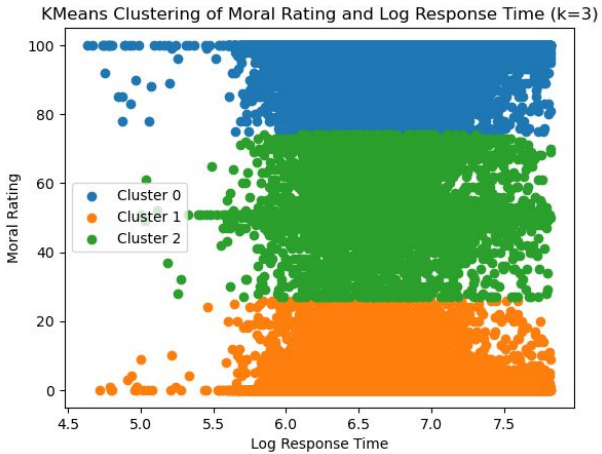


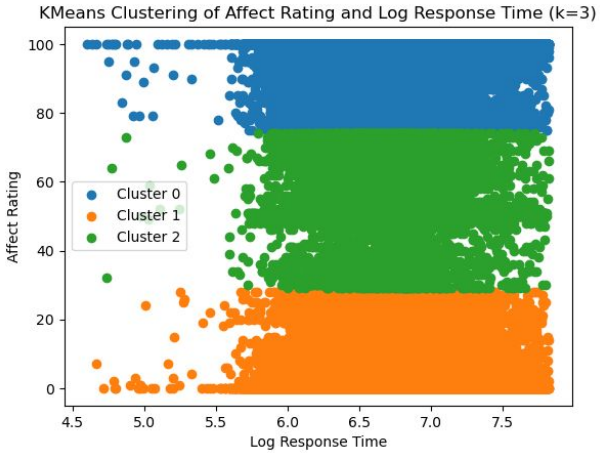
	R2	MAE
Ridge	0.027	396.68
RF	0.029	364.24
SVR	-0.009	371.82
XGBoost	-0.044	358.82

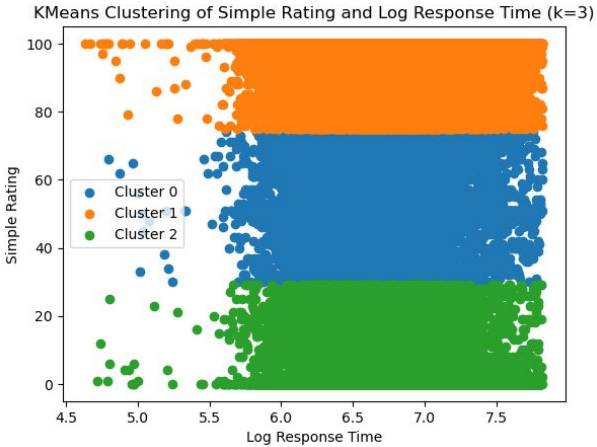


Are these the ‘right’ stimuli?:
Unsupervised ML Validation

Morally Good = the most pleasant and medium complex







Morally Bad = the least pleasant and least complex

Next Steps: MAAT Study 1

Optimize machine learning approaches

- Supervised: Predict specific stimuli, moral category of stimuli
- Unsupervised: More models to improve clustering

Develop a simulation-based power analysis

Examine asymmetries/differences in inferential and predictive relationships

- Compare participant-assigned moral valence to our assigned moral valence

Thank you for your time and interest.
Any questions?