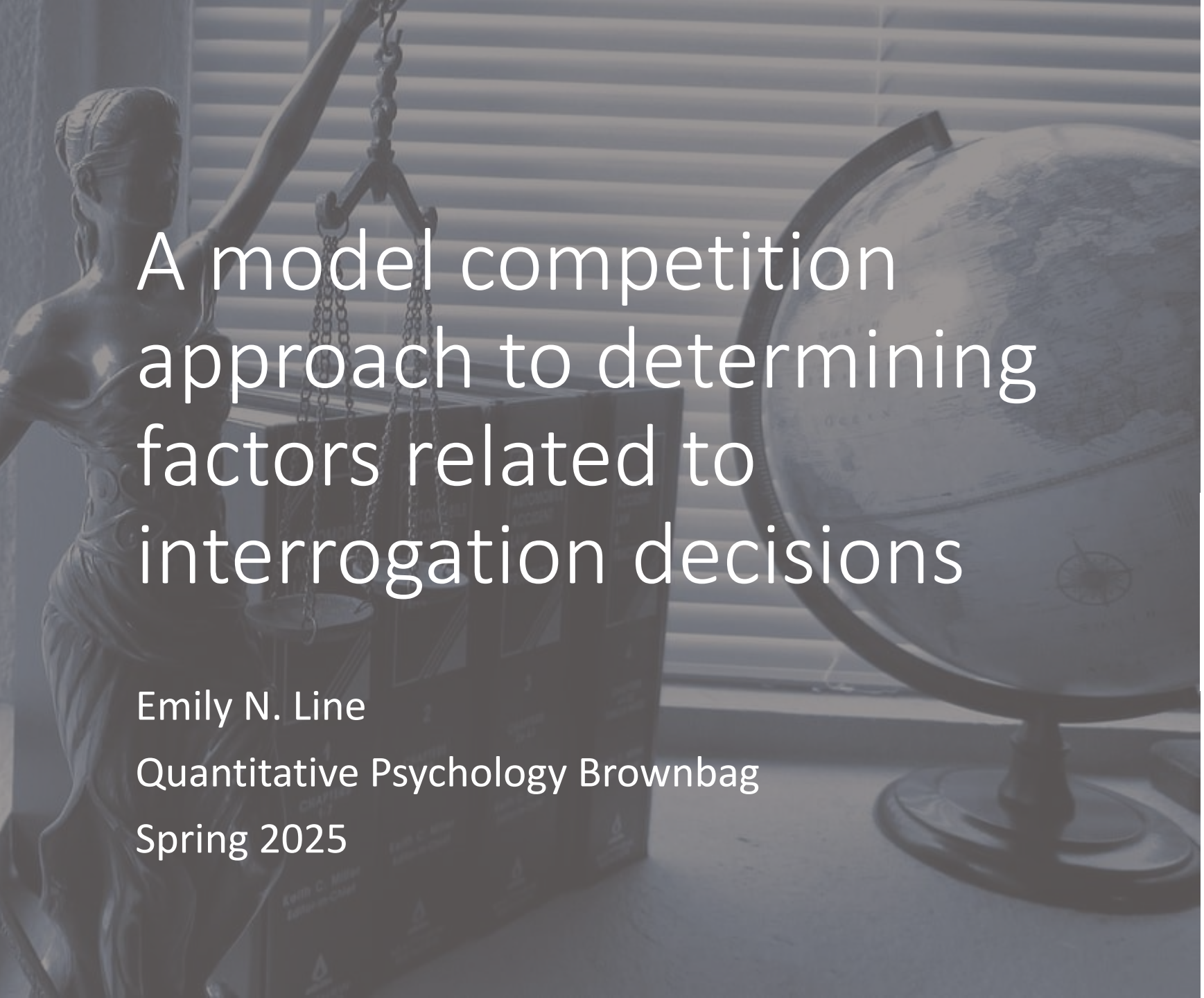


A model competition approach to determining factors related to interrogation decisions

Emily N. Line

Quantitative Psychology Brownbag

Spring 2025

Collaborators

Madison Harvey

Heather Price

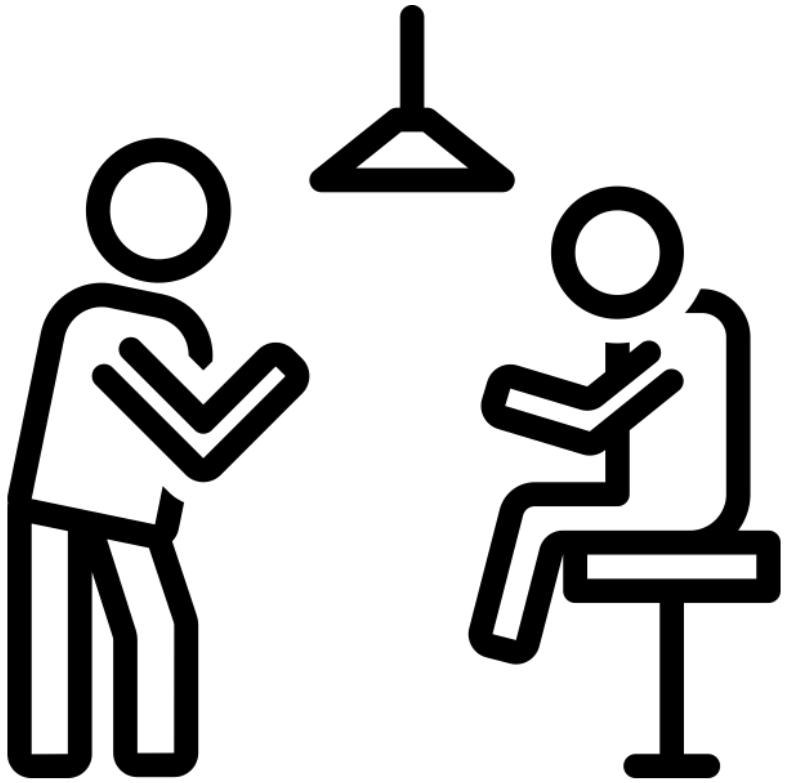
Dan Cavagnaro

Mike Regenwetter

Summer School 2022

Suspect Interrogation

- Suspect interrogations are an important part of the investigative process
  - Corroborate evidence
  - Assess suspect's behavior
  - Can significantly impact direction of investigation

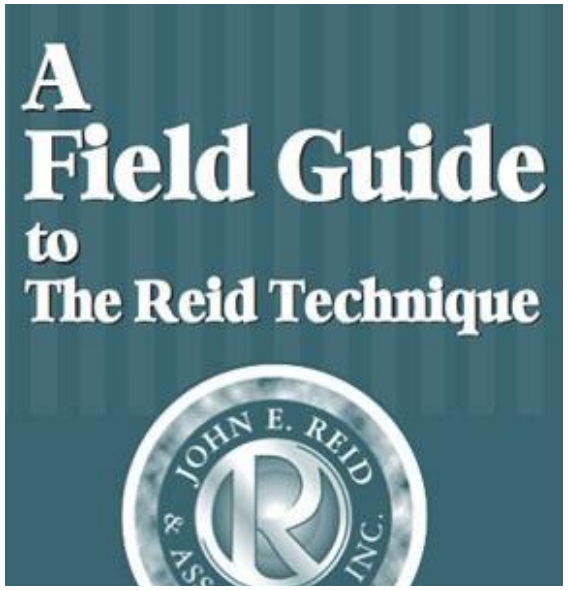


Created by Creative Mahira from Noun Project

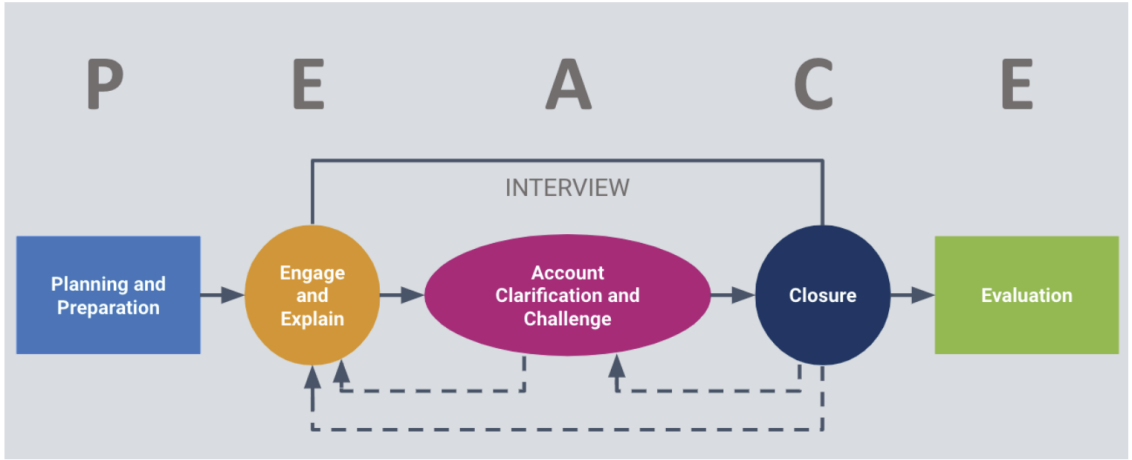
Interrogation Misconduct

- Interrogations can go wrong
- Amanda Knox
  - Falsely accused of murdering her roommate while abroad in Italy
  - Falsely confessed after a long, intense interrogation with aggressive police questioning
- 28% of wrongful exonerations involved false convictions (Innocence Project, 2025)
  - Average of 16 hours
  - Often no legal counsel present
  - Can involve police deception

Interrogation Decisions



- Interrogator's decisions and approach affects the likelihood of false confessions (Catlin et al., 2024)
- Accusatorial questioning (Reid technique; Inbau et al., 2013)
  - Start with presumption of guilt
  - Aim is to get suspect to confess
- Information-gathering approach
  - (PEACE technique; Bull & Rachlew, 2020)
  - Aim is to get credible information



Factors affecting interrogation decisions

- Presumption of guilt
  - Confirmation bias in seeking out evidence (O'Brien, 2009)
  - Asking more guilt-presumptive questions (Hill et al., 2010)
- Presence of tattoos (Brown et al., 2018)
  - More likely to be perceived as guilty
- Race
  - Mixed findings

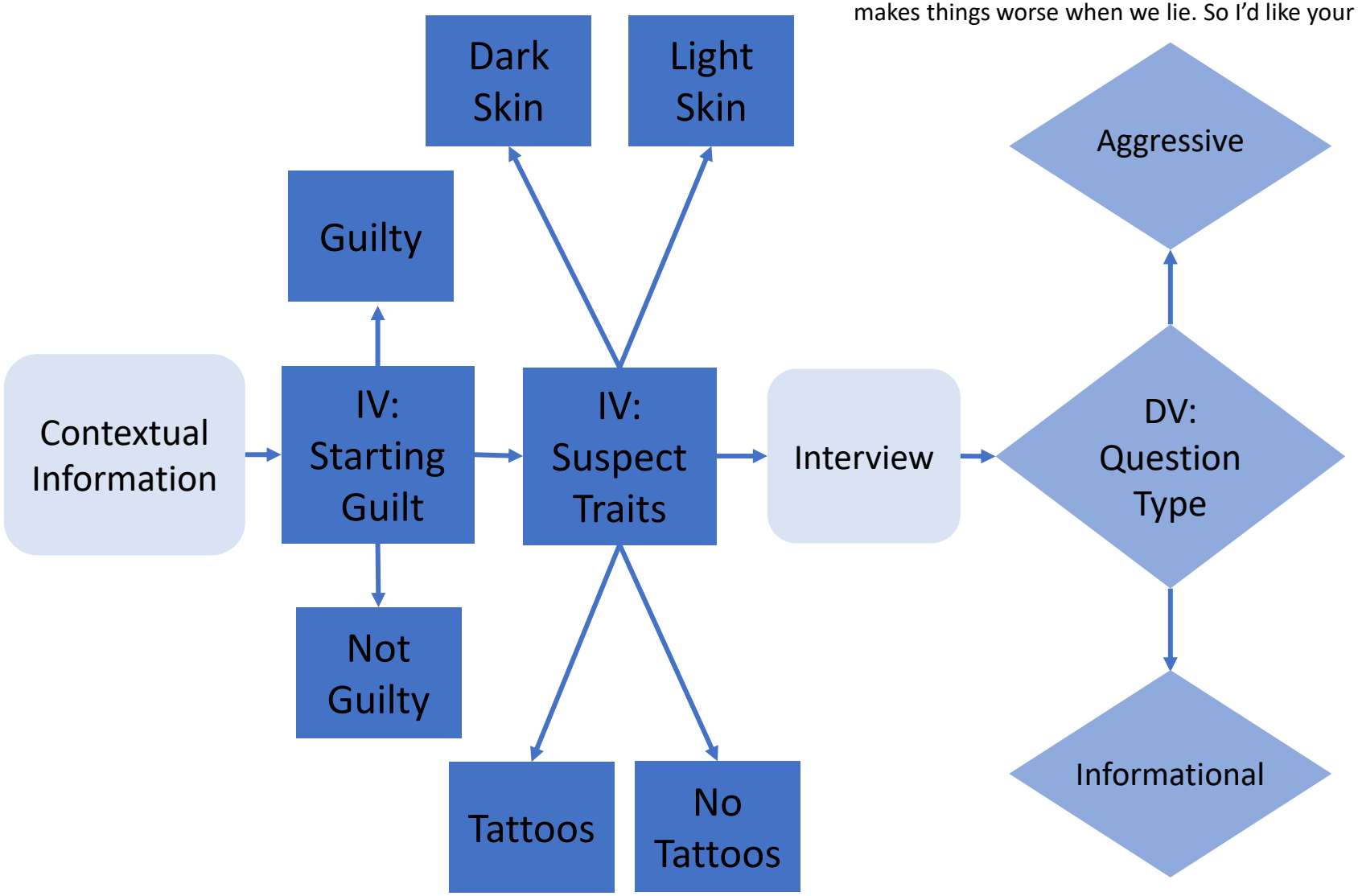
Race effects in legal contexts

- Conflicting findings in psychology and law regarding racial biases
  - Some studies find that (mock) jurors rate BIPOC individuals as less credible, and more likely to be guilty (Mitchell et al., 2005)
  - Other studies find the opposite effects (Estrada-Reynolds et al., 2022)
    - “Overcompensation” or “Reverse-race effect”
- Possibility of two sub-populations
  - Difficult to test using off-the-shelf statistical approaches

How do suspect traits and interrogators' pre-interview beliefs impact decision making?

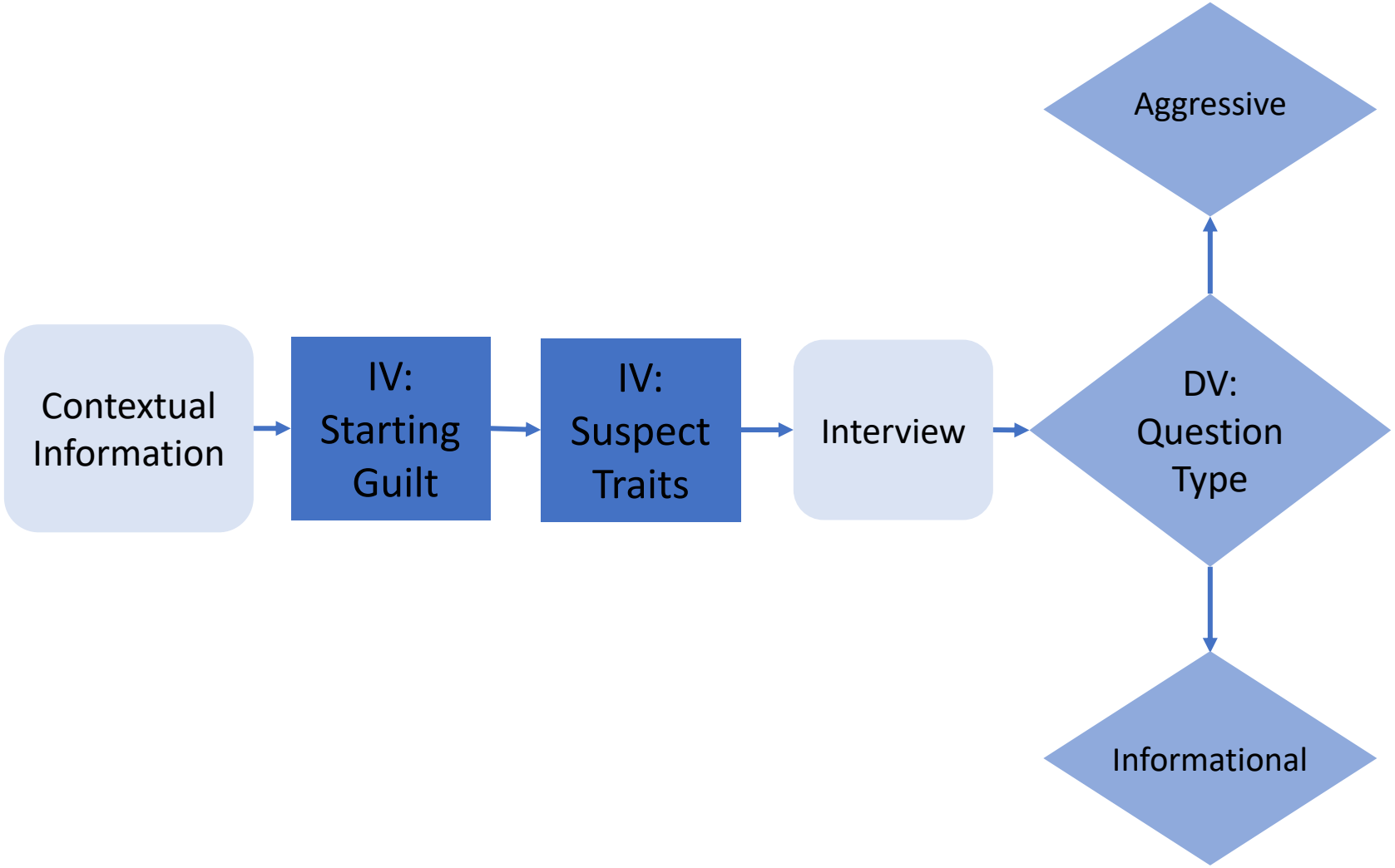
- 315 participants
- Mock interrogators complete a computer-based interview with an avatar

Well William, a neighbour saw your car there parked by her house that night. Your Toyota was seen at the house. See William, sometimes things happen accidentally, but it only makes things worse when we lie. So I'd like your side of things.



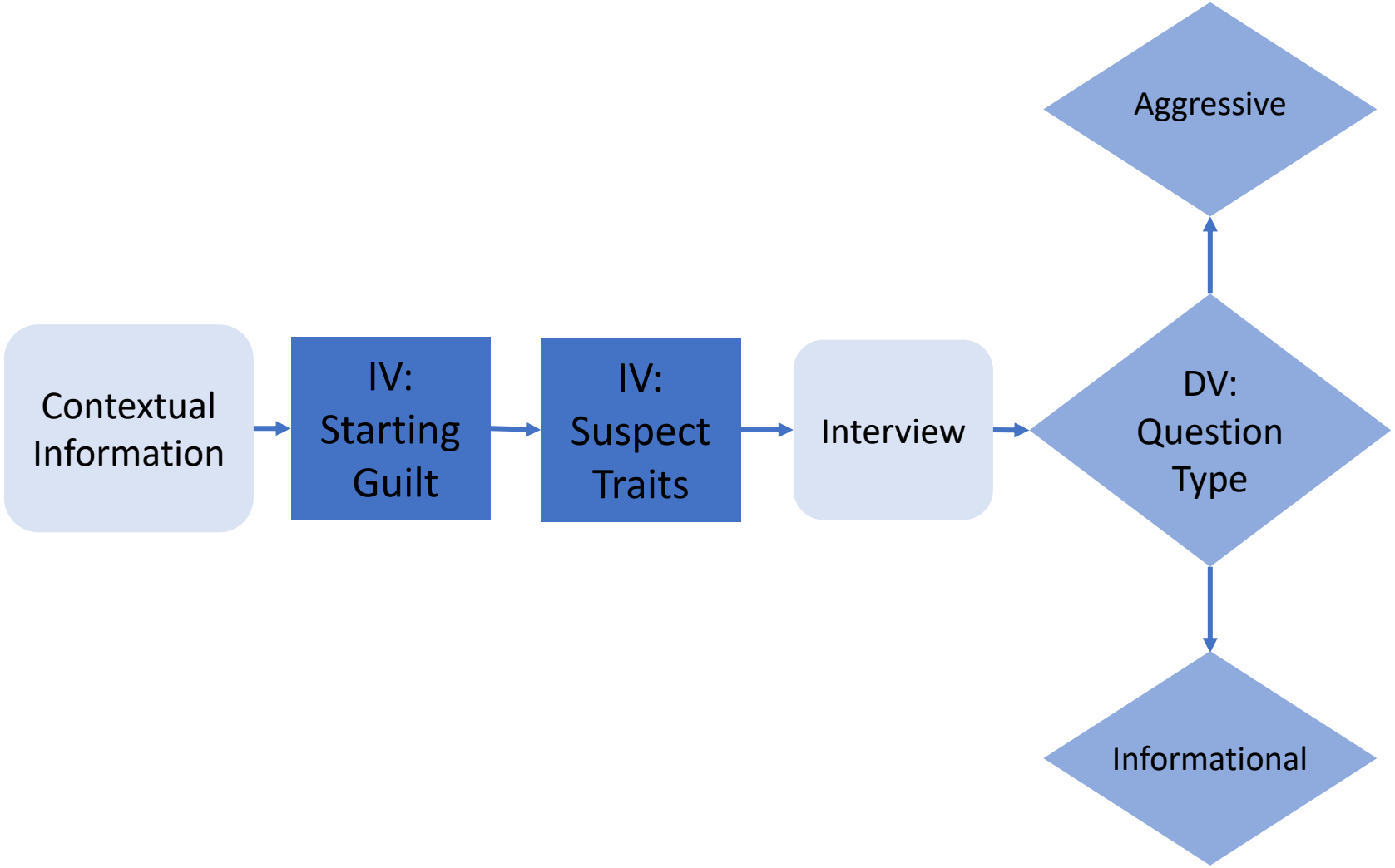
Well it's just that we had someone call in and suggesting, you know, to take a look at you. We just want to follow up on everything. Your vehicle, you drive a Toyota right, so we're just following up about the vehicles.

Well William, a neighbour saw your car there parked by her house that night. Your Toyota was seen at the house. See William, sometimes things happen accidentally, but it only makes things worse when we lie. So I'd like your side of things.



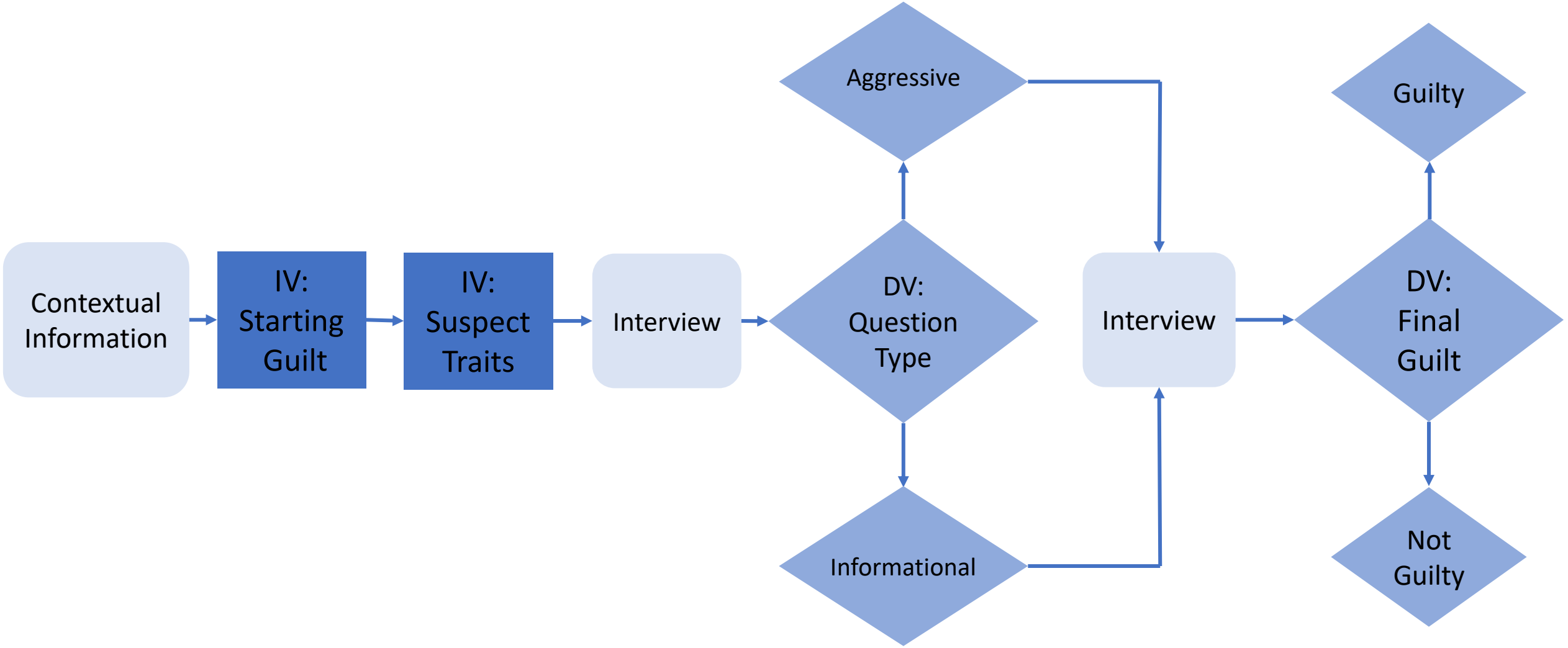
Well it's just that we had someone call in and suggesting, you know, to take a look at you. We just want to follow up on everything. Your vehicle, you drive a Toyota right, so we're just following up about the vehicles.

Well William, a neighbour saw your car there parked by her house that night. Your Toyota was seen at the house. See William, sometimes things happen accidentally, but it only makes things worse when we lie. So I'd like your side of things.

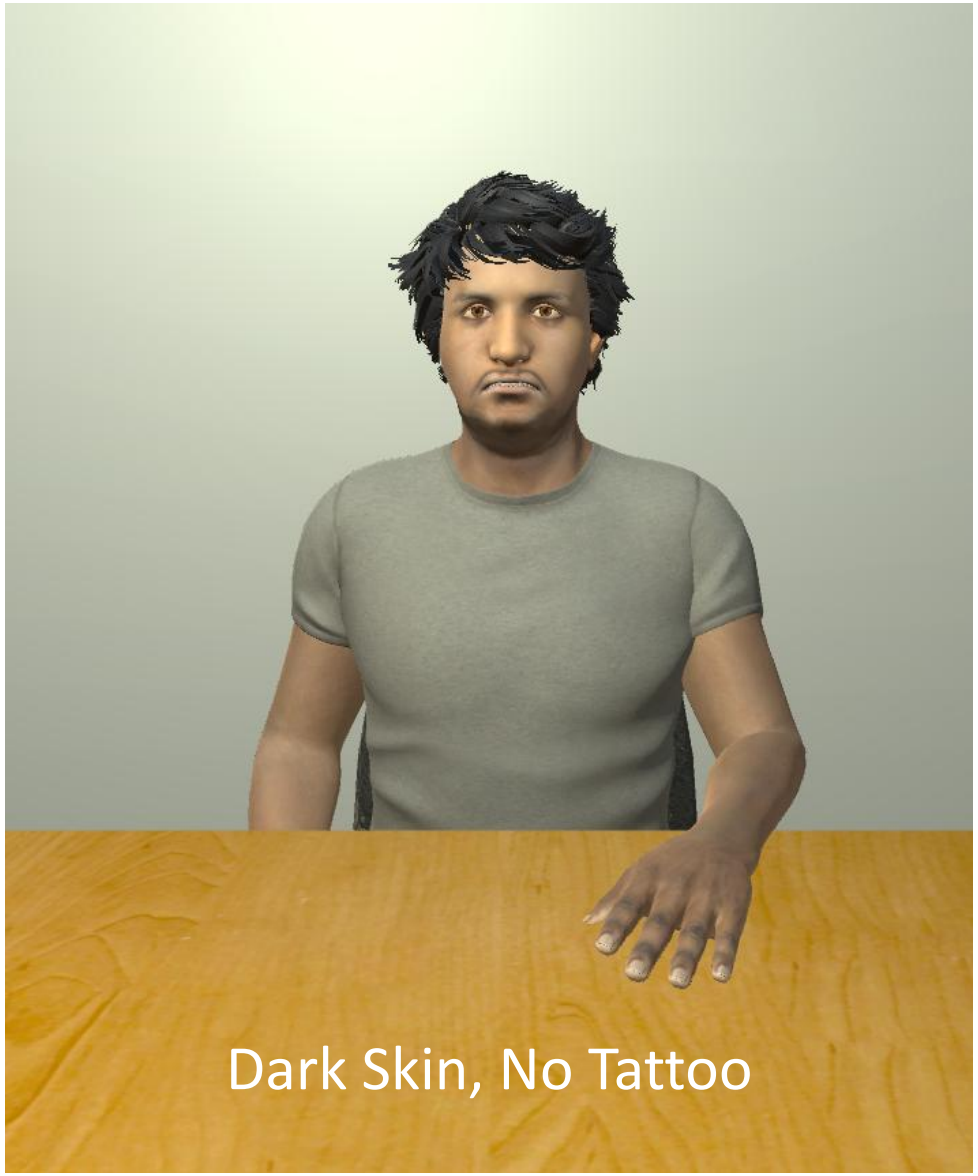


Well it's just that we had someone call in and suggesting, you know, to take a look at you. We just want to follow up on everything. Your vehicle, you drive a Toyota right, so we're just following up about the vehicles.

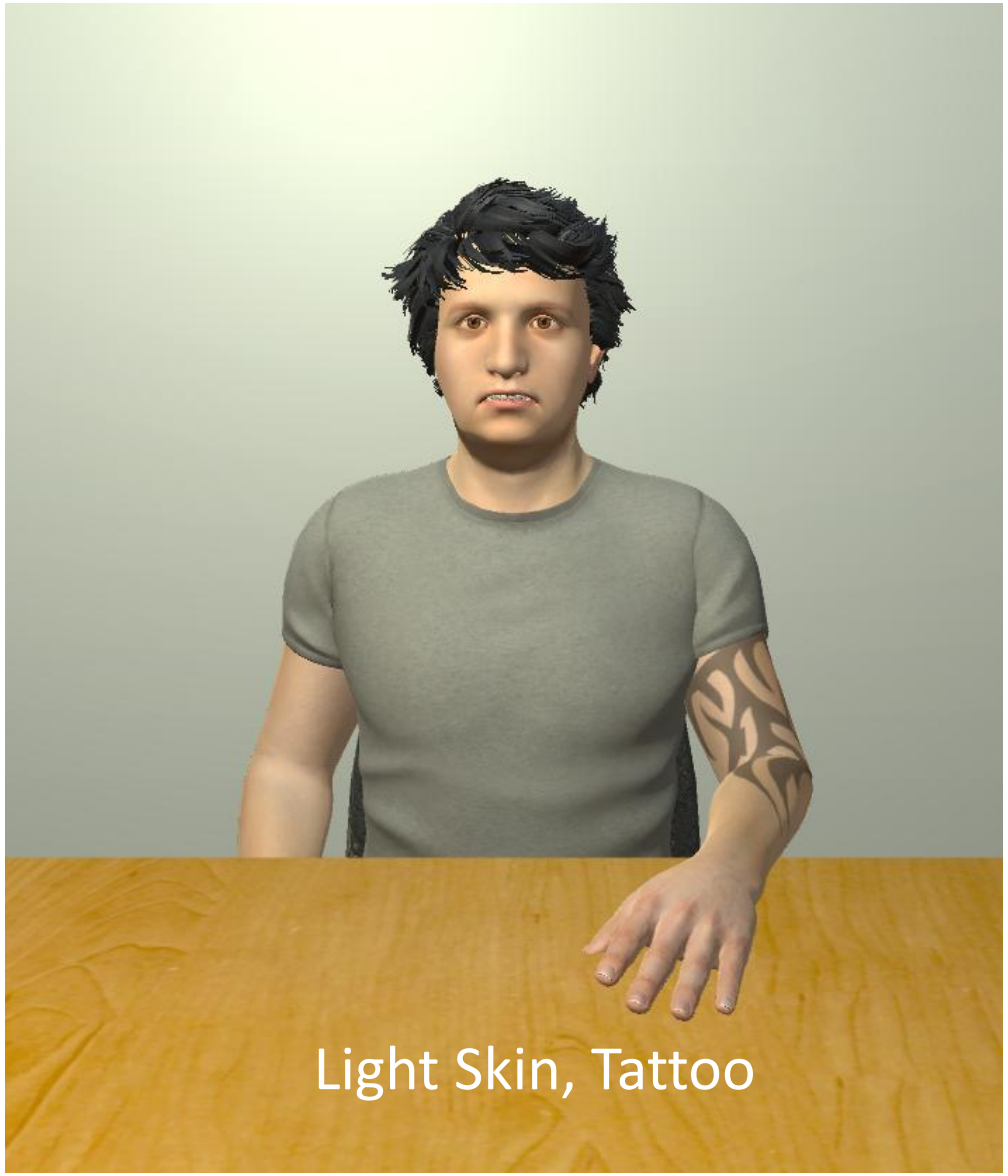
Well William, a neighbour saw your car there parked by her house that night. Your Toyota was seen at the house. See William, sometimes things happen accidentally, but it only makes things worse when we lie. So I'd like your side of things.



Well it's just that we had someone call in and suggesting, you know, to take a look at you. We just want to follow up on everything. Your vehicle, you drive a Toyota right, so we're just following up about the vehicles.



Dark Skin, No Tattoo



Light Skin, Tattoo

Guilt Presumption

- Not Guilty
- Guilty

Skin Color

- Light
- Dark

Tattoo Presence

- Tattoo
- No Tattoo

?

Question Type

- Aggressive
- Informational

Final Guilt

- Guilty
- Not Guilty

Model Space

- Eight groups:
  - Guilt Presumption (Not guilty/Guilty)
    x Skin Color (Light/Dark)
    x Tattoo Presence (Yes/No)
- Consider the probability of
  - Asking aggressive question
  - Giving a final judgment of guilty
- Translates to a 16-dimensional space

	Probability of Behavior			
	Guilty Pre		Innocent Pre	
	Dark	Light	Dark	Light
Tattoos	P_{DGT}	P_{LGT}	P_{DIT}	P_{LIT}
No Tattoos	P_{DGN}	P_{LGN}	P_{DIN}	P_{LIN}

Order-constrained inference

- Translate verbal hypotheses into mathematical order constraints
  - Comparing probabilities or sets of probabilities
- E.g., “Mock interrogators who view a suspect with dark skin are more likely to believe the suspect is guilty than interrogators who view a suspect with light skin.”

$$\mathbf{0} \leq \boldsymbol{P_L} \leq \boldsymbol{P_D} \leq \mathbf{1}$$

- Can use model competition to compare a set of competing hypotheses
  - Are people biased against suspects with dark skin, against suspects with light skin, or are there two sub-populations?

Hypotheses

- 22 hypotheses that consider possible effects of initial guilt presumption, tattoo presence, and race
- 6 mixture models that consider two sub-populations

See OSF:

General predictions across hypotheses

- People are more likely to ask aggressive questions/conclude guilt with suspects with dark skin compared to suspects with light skin
- People are more likely to ask aggressive questions/conclude guilt with suspects with light skin compared to suspects with dark skin
- Respondents are more likely to ask aggressive questions to suspects with tattoos than they are to suspects without tattoos
- Respondents who had a presumption of guilt will be more likely to ask aggressive questions than respondents with a presumption of not guilty

Skin Color
(bias against dark)

Skin Color
(bias against light)

Tattoos

Presumed Guilty

Additional variations between models

- Weighting of factors
  - E.g., Skin color > Presumption > Tattoos *(Hyp. 5 and 6)*
- Interactions
  - E.g., "Effect of skin color will be greater when participants have a presumption of guilt compared to a presumption of innocence" *(Hyp. 17 and 18)*
- Averaging vs. Overall "main effects"
  - E.g., $\{P_{LIT}+P_{LIN}+P_{DIN}+P_{DIT}\} \leq \{P_{DGT}+P_{DGN}+P_{LGN}+P_{LGT}\}$ *(Hyp. 15)*
  - *V.s.* $\{P_{LIT}, P_{LIN}, P_{DIN}, P_{DIT}\} \leq \{P_{DGT}, P_{DGN}, P_{LGN}, P_{LGT}\}$ *(Hyp. 16)*

Two example variations

Hyp 15: On average, participants are more likely to view a suspect as guilty when they had a pre-interrogation presumption of guilt:

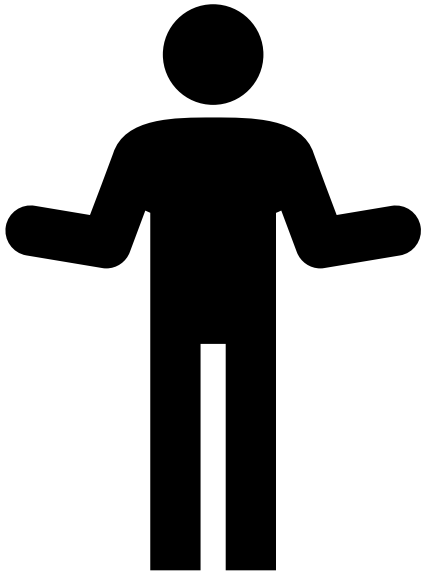
$$0 \leq \frac{P_{LIT}+P_{LIN}+P_{DIN}+P_{DIT}}{4} \leq \frac{P_{DGT}+P_{DGN}+P_{LGN}+P_{LGT}}{4} \leq 1$$

Hyp 16: Each group that presumed guilty will result in more guilty ratings than each group that presumed innocent:

$$0 \leq \{P_{LIT}, P_{LIN}, P_{DIN}, P_{DIT}\} \leq \{P_{DGT}, P_{DGN}, P_{LGN}, P_{LGT}\} \leq 1$$

Mixture Models

- Mixed findings regarding racial effects on legal judgments
- Could be indicative of two sub-populations:
  - One that exhibits biases against people of color
  - One that exhibits biases against white people (reverse-race effect)
- A mixture model allows for two sub-populations in the data
- Consider 6 models that would allow for these two potential sub-populations



The first 22 Hypotheses

Hypothesis	Skin Color (SC)	Tattoo (T)	Presumption (P)	Additional Components
1/2	D/L	✓	✓	None
3/4	D/L	✓	✓	P > SC > T
5/6	D/L	✓	✓	SC > P > T
7/8	D/L	X	✓	None
9/10	D/L	X	✓	P > SC
11/13	D/L	X	X	Average
12/14	D/L	X	X	Overall
15	X	X	✓	Average
16	X	X	✓	Overall

Hypothesis	Skin Color (SC)	Tattoo (T)	Presumption (P)	Additional Components
17/18	D/L	X	✓	Interaction (Larger SC effect when P = guilty)
19	D	✓	X	Interaction (T effect when SC = Dark)
20	D	✓	X	Interaction (Larger T effect when SC = Dark)
21	D	✓	✓	Interactions
22	D	✓	✓	Interactions

The first 22 Hypotheses

Hypothesis	Skin Color (SC)	Tattoo (T)	Presumption (P)	Additional Components
23 1/2	D/L	✓	✓	None
24 3/4	D/L	✓	✓	P > SC > T
25 5/6	D/L	✓	✓	SC > P > T
26 7/8	D/L	X	✓	None
27 9/10	D/L	X	✓	P > SC
11/13	D/L	X	X	Average
12/14	D/L	X	X	Overall
15	X	X	✓	Average
16	X	X	✓	Overall

Hypothesis	Skin Color (SC)	Tattoo (T)	Presumption (P)	Additional Components
28 17/18	D/L	X	✓	Interaction (Larger SC effect when P = guilty)
19	D	✓	X	Interaction (T effect when SC = Dark)
20	D	✓	X	Interaction (Larger T effect when SC = Dark)
21	D	✓	✓	Interactions
22	D	✓	✓	Interactions

Model Parsimony

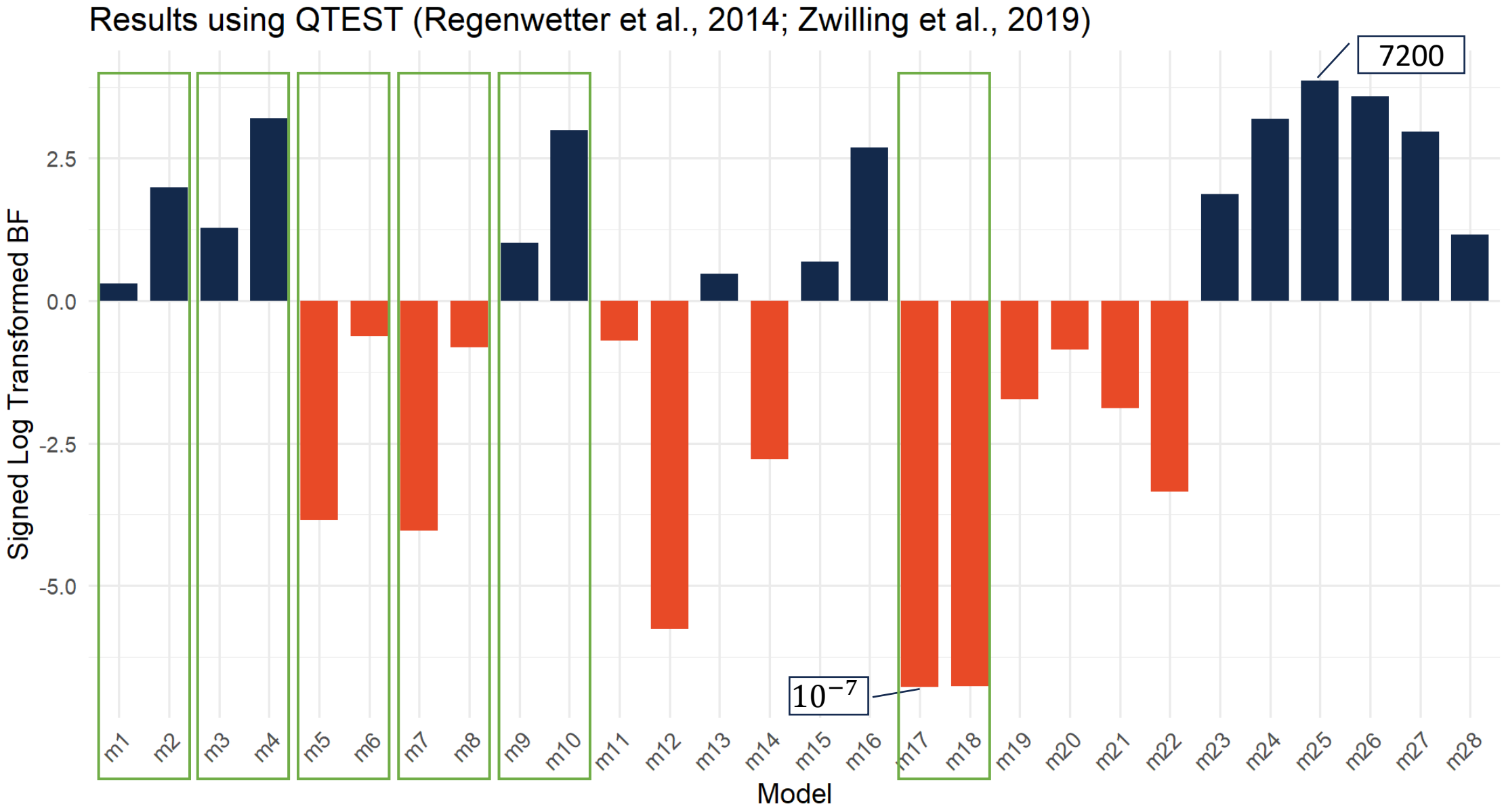
- Volume of model as a measure of model parsimony
- Maximum possible Bayes Factor (BF) against the unconstrained model is equal to $\frac{1}{volume}$

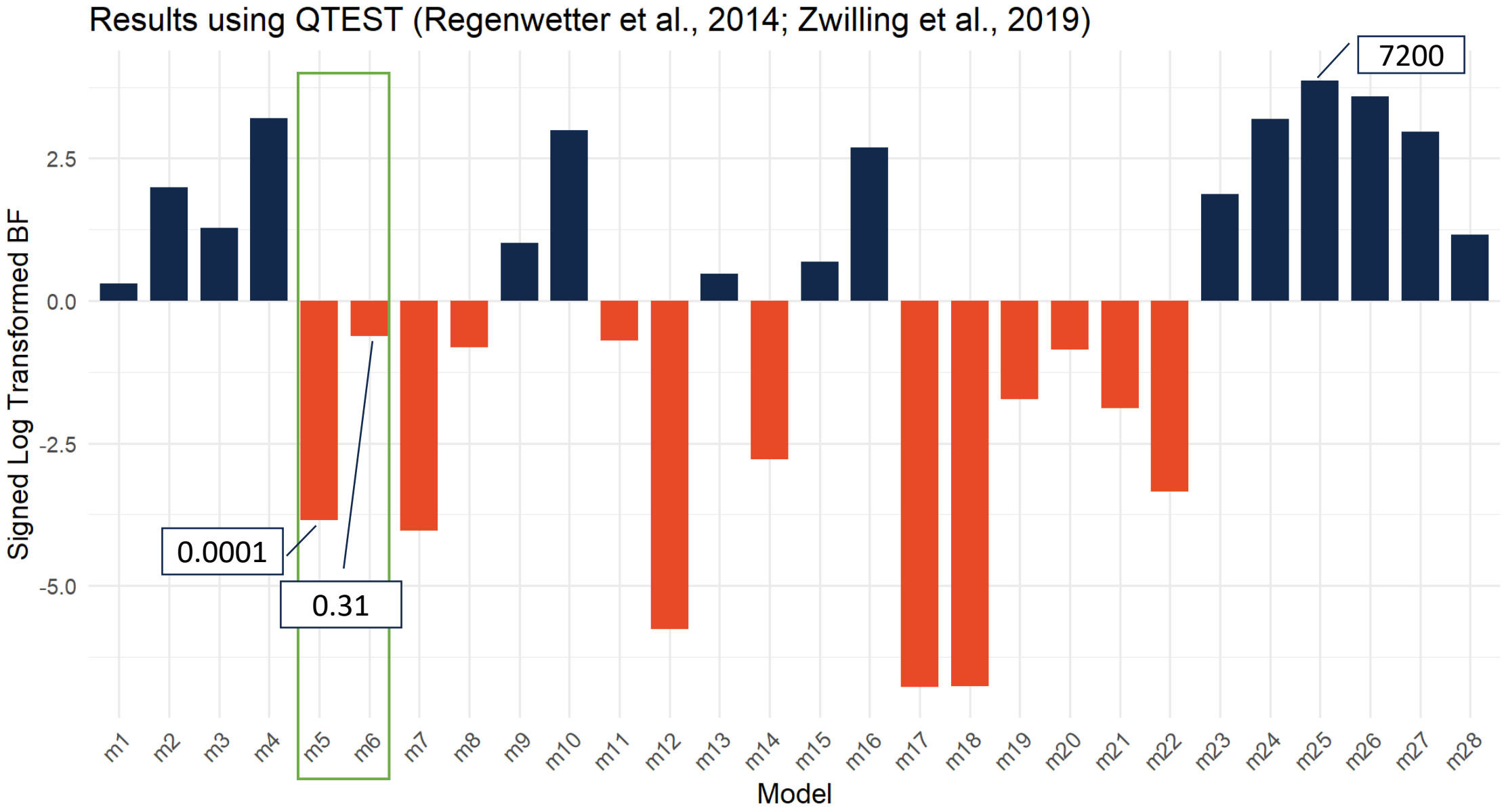
Model Parsimony		
Hypothesis	Volume	Max Possible BF
1	0.000001	1,000,000
2	0.000001	1,000,000
3	< 0.0000000016	62,500,000
4	< 0.0000000016	62,500,000
5	< 0.0000000016	62,500,000
6	< 0.0000000016	62,500,000
7	0.00000016	6,250,000
8	0.00000016	6,250,000
9	0.00000009	11,111,111
10	0.00000009	11,111,111
11	0.25	4
12	0.00020	5102
13	0.25	4
14	0.00021	4691

Model Parsimony		
Hypothesis	Volume	Max Possible BF
15	0.25	4
16	0.00020	4890
17	0.0080	125
18	0.0080	125
19	0.028	36
20	0.0076	132
21	0.0018	567
22	0.0070	142
23 (Mix 1 and 2)	0.000044	22,957
24 (Mix 3 and 4)	0.00000025	4,000,000
25 (Mix 5 and 6)	0.00000016	6,250,000
26 (Mix 7 and 8)	0.000045	22,277
27 (Mix 9 and 10)	0.00014	7062
28 (Mix 17 and 18)	0.72	1.38

Model Analyses

- First obtained Bayesian p-value as goodness-of-fit statistic
  - All models were a good fit ($p > .05$)
- Then obtained Bayes factor for each model
  - $$\frac{Evidence\ for\ Hypothesized\ Model}{Evidence\ for\ Baseline\ Model}$$
  - BF < 1 shows evidence against hypothesized model
  - BF > 1 shows evidence in support of hypothesized model


Results using QTEST (Regenwetter et al., 2014; Zwilling et al., 2019)
Signed Log Transformed BF
2.5
0.0
-2.5
-5.0
7200
10^{-7}
m1
m2
m3
m4
m5
m6
m7
m8
m9
m10
m11
m12
m13
m14
m15
m16
m17
m18
m19
m20
m21
m22
m23
m24
m25
m26
m27
m28
Model


Results using QTEST (Regenwetter et al., 2014; Zwilling et al., 2019)
Signed Log Transformed BF
2.5
0.0
-2.5
-5.0
0.0001
0.31
7200
m1
m2
m3
m4
m5
m6
m7
m8
m9
m10
m11
m12
m13
m14
m15
m16
m17
m18
m19
m20
m21
m22
m23
m24
m25
m26
m27
m28
Model

Mixture model – Hypotheses 5 and 6

Hypothesis 5 (6)

- Mock interrogators more likely to rate suspects with dark (light) skin as guilty compared to suspects with light (dark) skin
- Mock interrogators more likely to rate suspects with tattoos (vs. no tattoos) as guilty
- Mock interrogators with a presumption of guilt are more likely to rate suspects as guilty than interrogators with a presumption of innocence
- Skin color weighted more heavily than presumption of guilt/innocence, weighted more heavily than tattoos

Hyp 5 $P_{LIN} \leq P_{LIT} \leq P_{LGN} \leq P_{LGT} \leq P_{DIN} \leq P_{DIT} \leq P_{DGN} \leq P_{DGT}$

Hyp 6 $P_{DIN} \leq P_{DIT} \leq P_{DGN} \leq P_{DGT} \leq P_{LIN} \leq P_{LIT} \leq P_{LGN} \leq P_{LGT}$

Interrogation Decisions

- Evidence that there are two sub-populations regarding racial biases exhibited by participants
- Interrogators who presume the suspect is guilty are more likely to ask aggressive questions and judge suspect as guilty
  - “innocent until proven guilty”
- Physical characteristics beyond skin color also affect interrogators’ judgments
- Future research should investigate interventions or standardized methods of interrogation

Modeling Approach Conclusion

- Very precise, nuanced predictions using order constraints
- The conflicting race results in psychology and law research could be due to the presence of two sub-populations
- Assess possibility of sub-populations vs. homogeneous populations using model competition
- Novel contribution to the field of psychology and law

Thank you!

Emily N. Line

neuline2@illinois.edu

@emily-line

OSF Link:

Signed log transformation

$$\text{Transformed BF} = \begin{cases} -\log_{10}(1 + 1/BF), & \text{if } BF \leq 1 \\ \log_{10}(1 + BF), & \text{if } BF > 1 \end{cases}$$