澳門大學
UNIVERSIDADE DE MACAU
UNIVERSITY OF MACAU

GBA
YSFPS

Bot or not: How passenger tells apart AI and human drivers in the Turing test of automated driving?

Zhaoning LI[1, 2]

yc17319@umac.mo

@lizhn7

ME LAB

A.N.D Lab

[1]*Center for Brain and Mental Well-being, Department of Psychology, Sun Yat-sen University, Guangzhou, China,*
[2]*Center for Cognitive and Brain Science, Department of Psychology, University of Macau, Macau SAR, China*

1, 350, 000

Background

1, 350, 000*

Automated driving have the potential to increase road safety, as they can react faster than human drivers and are not subject to human errors.

* World Health Organization. (2018). Global status report on road safety 2018.

Background

Despite the potential benefits, there is no large scale deployment of autonomous cars (ACs) yet.

Existing literature has highlighted that the acceptance of the AC will increase if it drives in a human-like manner.

A variety of algorithms concern:

- *Human-like driving trajectories*
- *Human-like decision-making at intersections*
- *Human-like car following*
- *Human-like braking behaviour*
- *Human-like 'gap acceptance' behaviour at crossings*
- *Human-like 'peeking' when approaching road junctions*
- *Human-like cost function*
- *Human-like driving policies in collision avoidance and merging*

Teaching ACs about human-like driving from the algorithmic perspective

Background

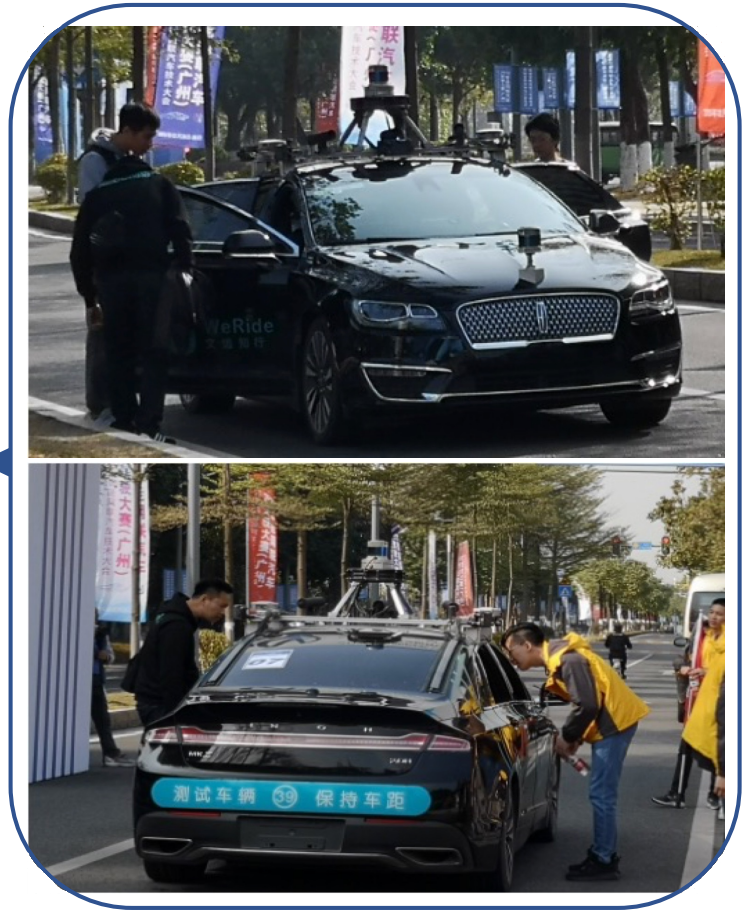
Despite the potential benefits, there is no large scale deployment of autonomous cars (ACs) yet.

Existing literature has highlighted that the acceptance of the AC will increase if it drives in a human-like manner.

However, literature presents no human-subject research focusing on passengers in a natural environment that examines whether the AC should behave in a human-like manner.

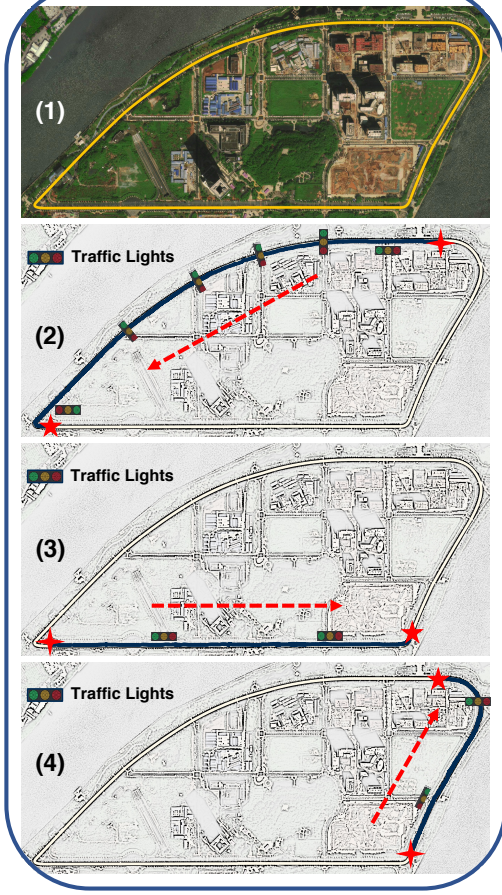
How to offer naturalistic experiences from a passenger's seat perspective to measure the people's acceptance of ACs?

The Turing test of automated driving



A

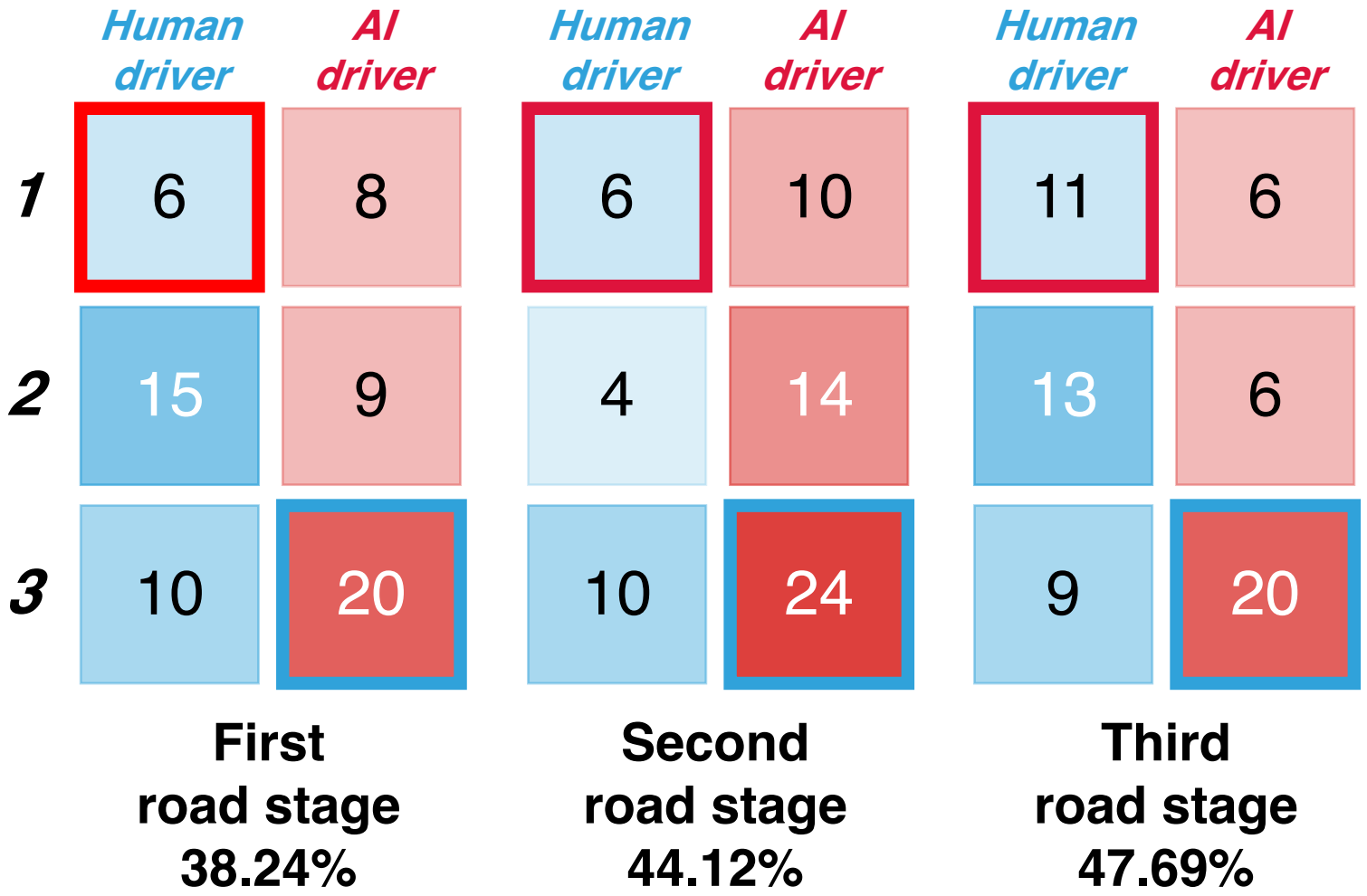
B



C

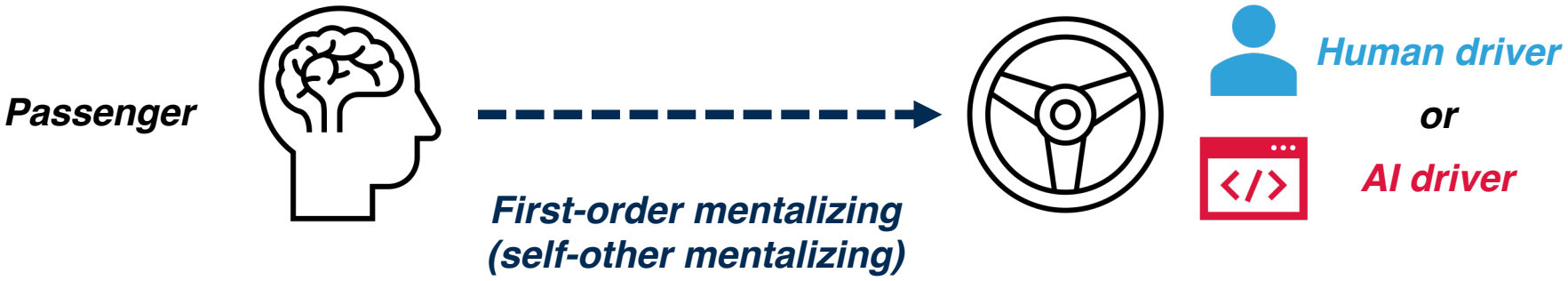
Results of the Turing test

Confusion matrix of three stages for the results in the Turing test



How do human passengers choose in the Turing test of automated driving?

How do human passengers choose?



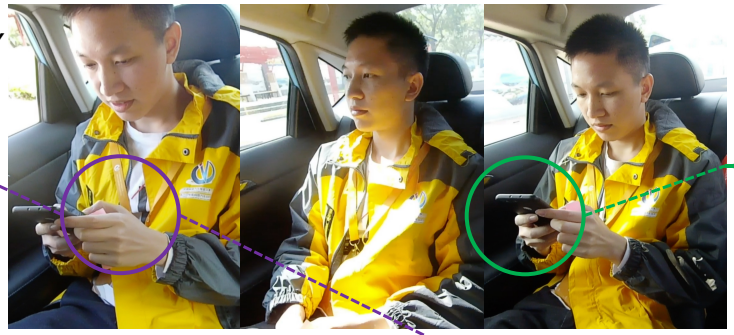
Choice behaviour $\rightarrow B = f(P, E)$

Kurt Lewin, (1936)

Passenger **Driving environment**

How do human passengers choose?

A. Participant data



Pre-study baseline:

DES-IV

Post-stage:

Response

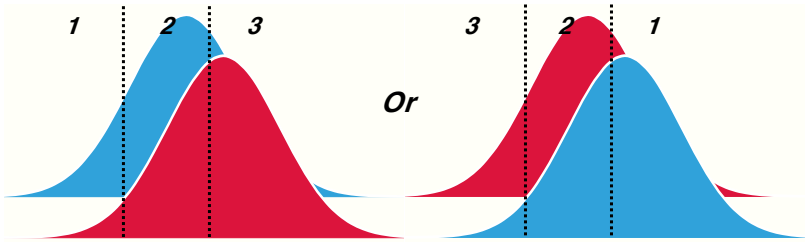
Safety and comfort

DES-IV

Other feelings

B. Signal detection theory

Unlikely (1) / somewhat likely (2) / very likely (3) to be driven by the AI driver



Stimuli: Human driver and AI driver

Signal strength

1 / 2 / 3 ≈ { [( ), ( )], / }

C. Affective variability

( ): Pre-study baseline vector

( ): Post-stage vector

Dissimilarity measures

Or 1 - Cosine similarity

Or Euclidean distance

Or Manhattan distance

Or Word mover's distance

Or Word rotator's distance

D. Transformation

较强烈快乐 Enjoyment (3/4)

较强烈兴趣 Interest (3/4)

较轻微惊奇 Surprise (2/4)

一点也没有恐惧 Fear (1/4)

一点也没有紧张 Tension (1/4)

较强烈满意 Satisfaction (3/4)

过红绿灯时停车较急促。The car stopped more quickly at traffic lights.

Pre-trained language models

Embedding

Sentence level

较 强 烈 快 乐 … …

Or

Document level

较强烈快乐
较强烈兴趣
较轻微惊奇
一点也没有恐惧
一点也没有紧张
较强烈满意
过红绿灯时…

Pooling

Max Mean Min

Or Or

Or Max-mean-min

Or Max-mean

Or Max-min

Or Mean-min

Whitening and dimensionality reduction

Transformed vector

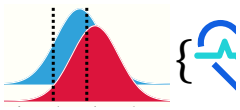
How do human passengers choose?

A. Parti

Pre-study baseline:

DES-IV

1 / 2 / 3 ≈



较强烈快乐 Enjoyment (3/4)

较强烈兴趣 Interest (3/4)

较轻微惊奇 Surprise (2/4)

一点也没有恐惧 Fear (1/4)

一点也没有紧张 Tension (1/4)

较强烈满意 Satisfaction (3/4)

过红绿灯时停车较急促。The car stopped more quickly at traffic lights.

Pre-trained language models

KeAi CHINESE ROOTS GLOBAL IMPACT

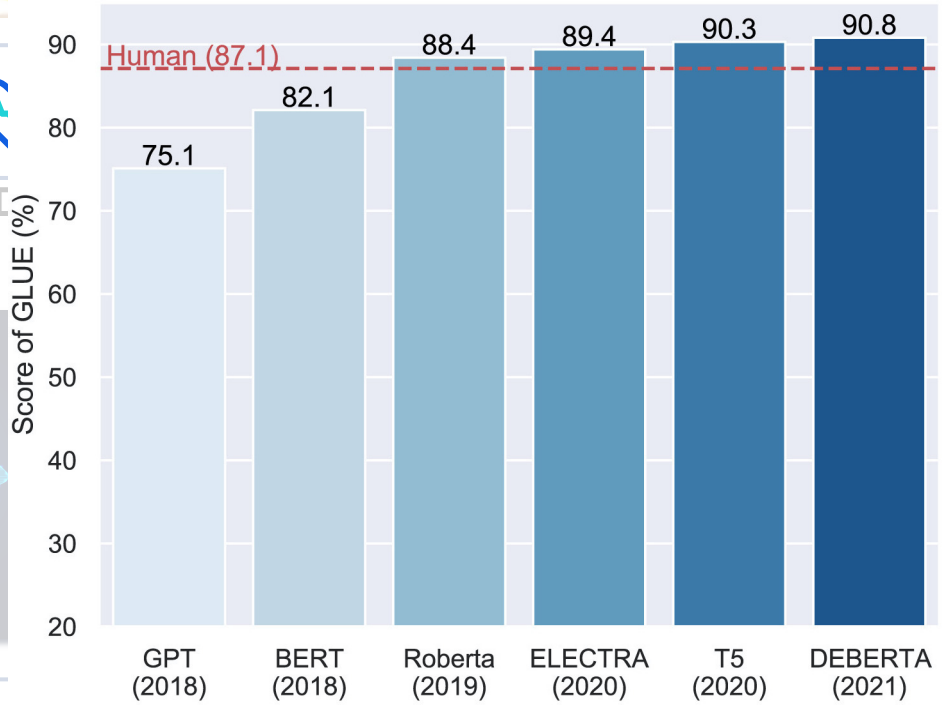
AI Open

Available online 26 August 2021

In Press, Journal Pre-proof

Pre-Trained Models: Past, Present and Future

Xu Han [1] ✉*, Zhengyan Zhang [1,*] ✉, Ning Ding [1,*] ✉, Yuxian Gu [1,*] ✉, Xiao Liu [1,*] ✉, Yuqi Huo [2,*] ✉, Jiezhong Qiu [1] ✉, Liang Zhang [2] ✉, Wentao Han [1,†] ✉, Minlie Huang [1,†] ✉, Qin Jin [2,†] ✉, Yanyan Lan [4,†] ✉, Yang Liu [1,4,†] ✉, Zhiyuan Liu [1,†] ✉, Zhiwu Lu [3,†] ✉, Xipeng Qiu [5,†] ✉, Ruihua Song [3,†] ✉, Jie Tang [1,†] ✉ ... Jun Zhu [1,†] ✉



detection theory

newhat likely (2) / very likely (3)

driven by the AI driver

Signal strength

Affective variability

Pre-study baseline vector

Post-stage vector

Dissimilarity measures

Or 1 - Cosine similarity

Or Euclidean distance

Or Manhattan distance

Or Word mover's distance

Or Word rotator's distance

Results of the computational models

Comparison on the Outer Loop Cross Validation of Nested LOOCV with Baselines

(a) Evaluation results on the first stage.

Models	*ACC*	*P*	*R*	*F1*	ρ
Baselines					
Random	33.27	33.21	33.25	32.27	0.07
Probability	36.14	33.24	33.26	33.00	-0.68
God	38.24	24.47	36.51	28.79	14.91
SDT-AV					
Original	33.82	27.36	28.21	27.09	16.31
PLM-tf (AA)	51.47	50.71	**51.11**	50.30	**56.25*****
PLM-tf (AA+OF)	**54.41**	**50.94**	50.08	**50.37**	38.96**

Results of the computational models

Comparison on the Outer Loop Cross Validation of Nested LOOCV with Baselines

(a) Evaluation results on the first stage.

(b) Evaluation results on the second stage.

Models	*ACC*	*P*	*R*	*F1*	ρ
Baselines					
Random	33.35	33.37	33.36	32.15	0.15
Probability	37.71	33.55	33.58	33.32	0.25
God	44.12	26.67	36.03	30.62	3.94
SDT-AV					
Original	45.59	41.20	37.19	36.92	15.43
PLM-tf (AA)	57.35	56.65	53.80	54.59	36.46**
PLM-tf (AA+OF)	**63.24**	**59.74**	**56.62**	**57.48**	**41.20*****

Results of the computational models

Comparison on the Outer Loop Cross Validation of Nested LOOCV with Baselines

(a) Evaluation results on the first stage.

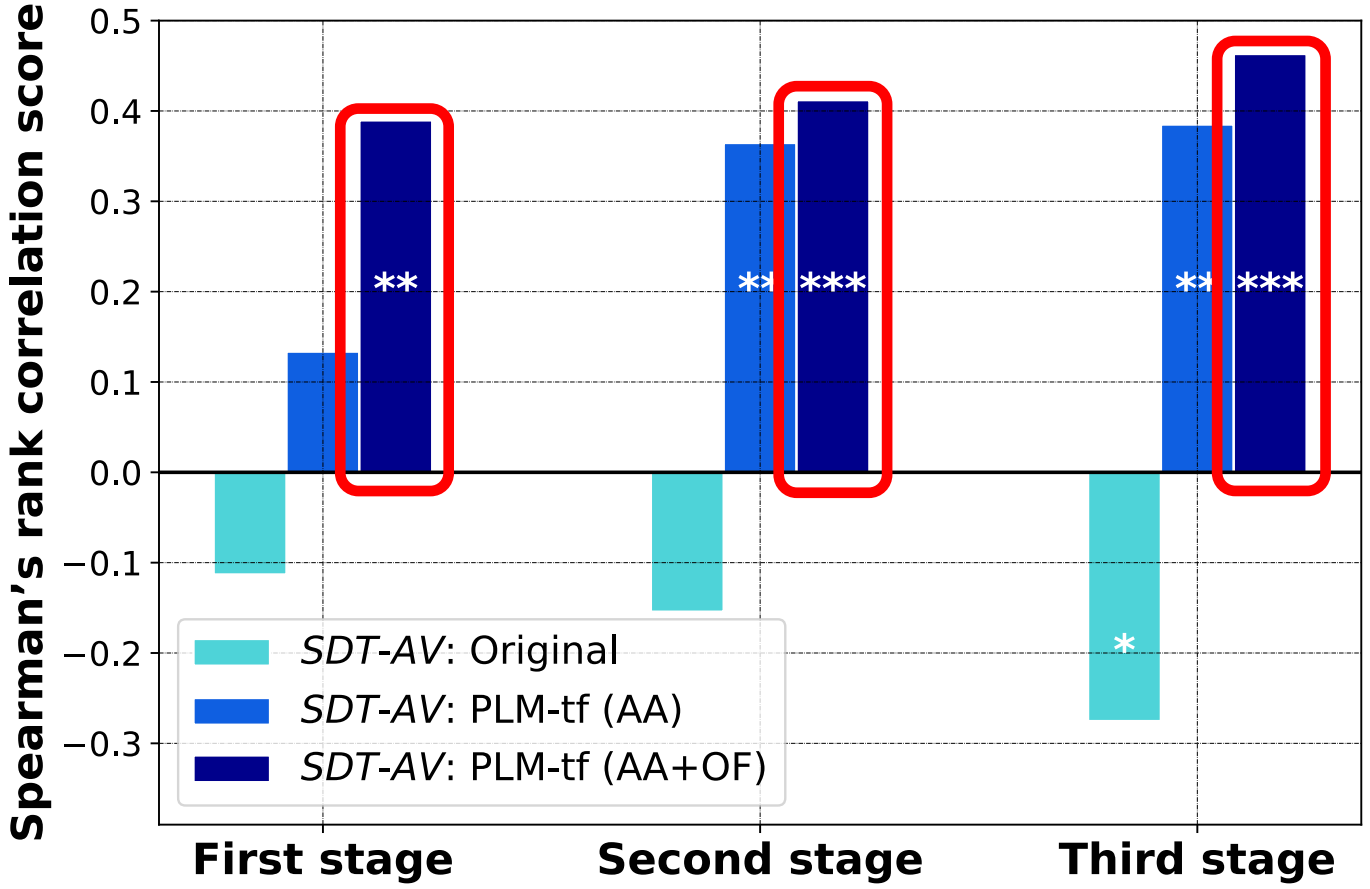
(b) Evaluation results on the second stage.

(c) Evaluation results on the third stage.

Models	*ACC*	*P*	*R*	*F1*	ρ
Baselines					
Random	33.40	33.34	33.39	32.66	-0.58
Probability	35.14	33.13	33.16	32.87	-0.15
God	47.69	31.94	44.56	36.52	31.68*
SDT-AV					
Original	53.85	48.84	45.62	45.42	27.54*
PLM-tf (AA)	52.31	49.65	49.81	49.67	38.50**
PLM-tf (AA+OF)	**55.38**	**51.81**	**51.56**	**51.67**	**46.31*****

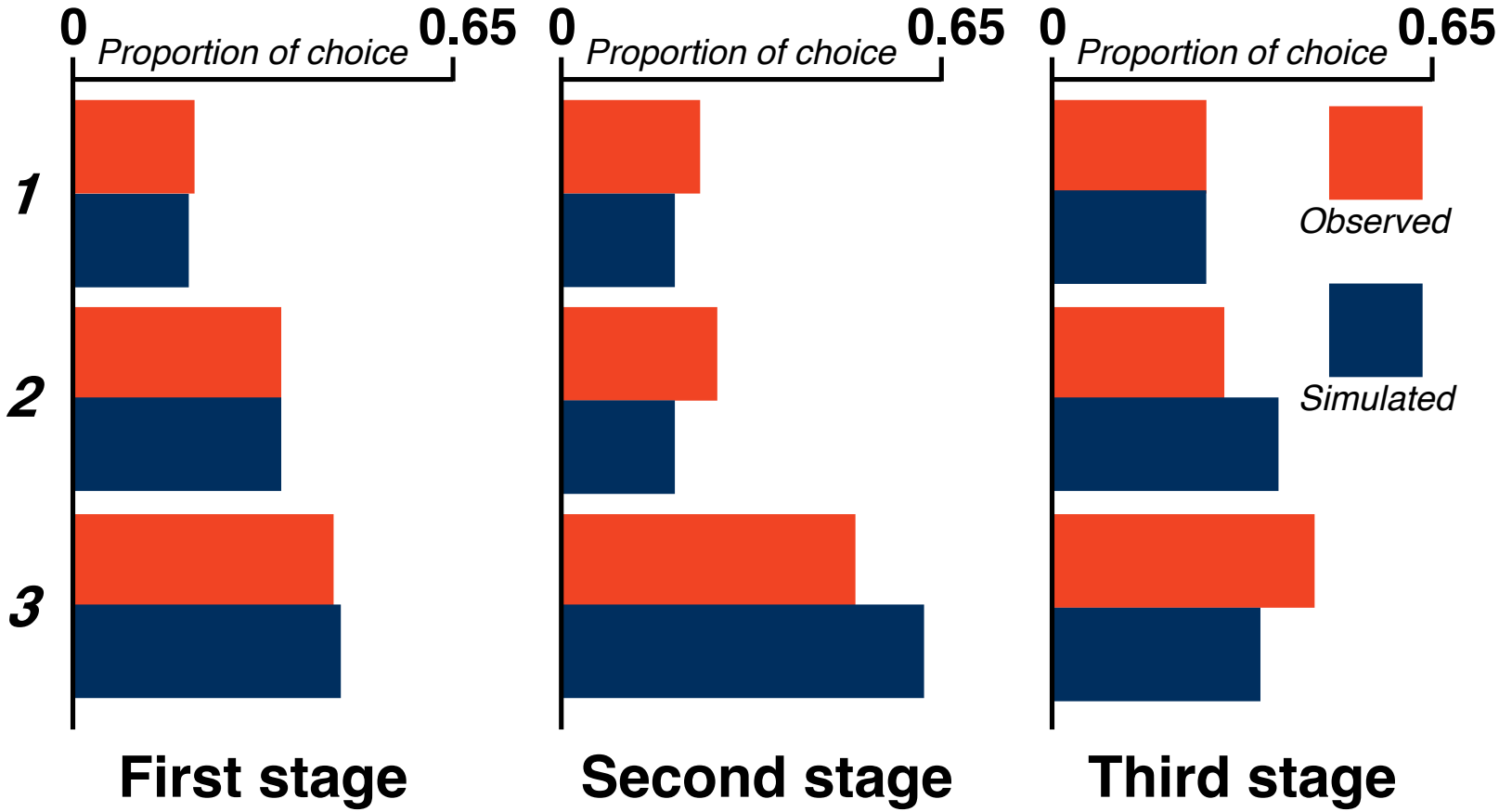
Correlations between choice of preference and affective variability

Comparison of the Spearman's rank correlation score between the gold labels and the magnitude of affective variability



Ordinal logistic regression analysis of model simulations

Comparison of the proportion of choices between model simulations (blue) and empirically observed choices (red)



Ordinal logistic regression analysis of model simulations

(a) Results of OLR predicting simulated labels on the first stage.

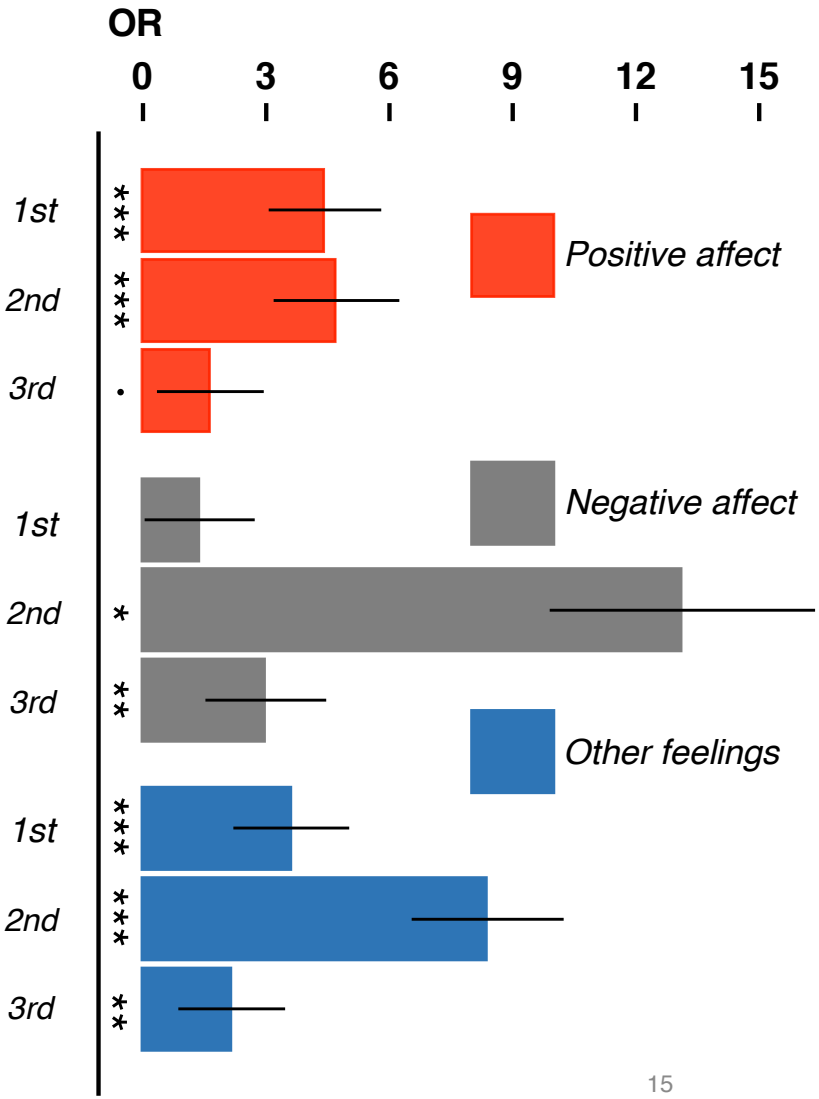
Coeff.	β (*SE*)	*t* Value	OR (*95% CI*)	*p* Value
I (1|2)	**-2.31 (0.47)**	**-4.92**		**<.0001*** **
I (2|3)	0.40 (0.31)	1.26		.208
PA	**1.49 (0.32)**	**4.66**	**4.42 (2.47-8.72)**	**<.0001*** **
NA	0.31 (0.29)	1.08	1.37 (0.78-2.47)	.28
OF	**1.29 (0.34)**	**3.74**	**3.62 (1.93-7.54)**	**<.001*** **

(b) Results of OLR predicting simulated labels on the second stage.

Coeff.	β (*SE*)	*t* Value	OR (*95% CI*)	*p* Value
I (1|2)	**-3.85 (0.85)**	**-4.55**		**<.0001*** **
I (2|3)	**-1.72 (0.65)**	**-2.67**		**.008****
PA	**1.55 (0.42)**	**3.65**	**4.70 (2.23-12.11)**	**<.001*** **
NA	**2.57 (1.17)**	**2.19**	**13.11 (2.10-226.37)**	**.028***
OF	**2.12 (0.61)**	**3.47**	**8.37 (3.04-35.96)**	**<.001*** **

(c) Results of OLR predicting simulated labels on the third stage.

Coeff.	β (*SE*)	*t* Value	OR (*95% CI*)	*p* Value
I (1|2)	**-1.35 (0.33)**	**-4.04**		**<.0001*** **
I (2|3)	**0.80 (0.30)**	**2.63**		**.009****
PA	0.49 (0.26)	1.86	1.63 (0.98-2.78)	.062
NA	**1.09 (0.38)**	**2.83**	**2.97 (1.56-7.14)**	**.005****
OF	**0.77 (0.26)**	**2.93**	**2.15 (1.31-3.69)**	**.003****



Summary

We conduct a Turing test of automated driving based on 69 passengers' feedback in a real scenario, and test results show that SAE Level 4 ACs could pass the Turing test with accuracy no more than 50%.

On this basis, we propose a model combining SDT with AV (transformed by PLMs) to predict the passenger's choice behaviour in the Turing test. This is, to the best of our knowledge, the first computational model which provides a mechanistic understanding underlying passengers' mentalizing process.

Extensive experimental results and further analysis show that the the greater AV that passengers have, the more likely they identify the driver as the AI algorithm. These findings suggest that future automated driving should improve the affective stability of passengers.

Thanks for your attendance!