



# Integrating social neuroscience into human-machine mutual behavioral understanding for autonomous driving

Yingji Xia,<sup>1</sup> Hui Chen,<sup>2</sup> and Xiqun Chen<sup>1,3,4,\*</sup>

<sup>1</sup>Institute of Intelligent Transportation Systems, College of Civil Engineering and Architecture, Zhejiang University, Hangzhou 310058, China

<sup>2</sup>Department of Psychology and Behavioral Sciences, Zhejiang University, Hangzhou 310058, China

<sup>3</sup>Zhejiang University/University of Illinois Urbana-Champaign (ZJU-UIUC) Institute, Haining 314400, China

<sup>4</sup>Zhejiang Provincial Engineering Research Center for Intelligent Transportation, Hangzhou 310058, China

\*Correspondence: [chenxiqun@zju.edu.cn](mailto:chenxiqun@zju.edu.cn)

Received: April 17, 2023; Accepted: May 28, 2023; Published Online: May 30, 2023; <https://doi.org/10.1016/j.xinn.2023.100455>

© 2023 The Author(s). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Citation: Xia Y., Chen H., and Chen X. (2023). Integrating social neuroscience into human-machine mutual behavioral understanding for autonomous driving. *The Innovation* 4(4), 100455.

Autonomous vehicles (AVs) are advertised to free human drivers, providing a safer and more efficient transport mode. After decades of extensive investment and invention, various types of AVs have been unveiled, but they are still restricted to limited application scenarios because of potential safety concerns. Despite rare sensing or detection failures from corner cases, one of the significant concerns primarily questions whether AVs would interact appropriately with surrounding human-driven vehicles on public roads. Particularly, the lack of approaches to human-like mutual driving understanding challenges the driving safety situation of AVs and human drivers. That is, human informal driving rules and implicit driving interactions cannot be understood explicitly by AVs, and human drivers can hardly accommodate the stilted or inconsistent driving behaviors generated by AVs because of distinctive “driving behavior understanding” mechanisms.

Due to the current small market penetration rate of AVs, they will inevitably share roads with human-driven vehicles for a long term. In other words, interactions between human-driven vehicles and AVs may challenge road safety for a prolonged period. Moreover, the opacity of AV decision-making algorithms brings psychological roadblocks to human drivers’ trust, which prevents public acceptance and affects the adoption of AVs. Therefore, barriers to mutual driving understanding underline the urgent need for investigation.

Because driving interaction could be formulated as a cooperative task, both vehicle types require mutual understanding and cooperative road sharing. In this Editorial, we point out the current research bottleneck in AV driving understanding and gain insights into the integration with state-of-the-art social neuroscience research, such as interbrain synchrony or social neuromorphic computing, to achieve human-machine mutual driving understanding.

## BOTTLENECKS IN AV DRIVING UNDERSTANDING RESEARCH

Recent research advances confirm that state-of-the-art AV research is dominated by various data-driven models, such as deep learning models and their variants. Under this technology trend, most AV modules have been facilitated from driving scene understanding tasks to end-to-end autonomous route planning frameworks. Admittedly, AVs informed by these data-driven models function well under restricted laboratory environments or free-flow traffic conditions. However, the black-box nature of these data-driven models brings shortcomings and research bottlenecks for AVs to achieve mutual or even unidirectional driving understanding.

Generally, because data-driven models learn from data rather than explicit rules or knowledge, they are believed to have inherent reproducibility and interpretability bottlenecks. A plausible standpoint is that data-driven neural networks use node gradients and weights across different layers to fit all kinds of correlations between input data and training targets rather than proposing causal reasoning or mechanism-level analysis. Because proposed fittings in these black boxes lack physical meaning, the model may put out outliers, and confounding model variables are difficult to troubleshoot, which is unacceptable in safety-critical applications like driving.

The bottlenecks of data-driven models are magnified when we focus on their employment in AV’s driving understanding tasks. From the model’s perspective, the “state space” of driving behaviors and interactions is staggeringly vast, requiring massive recordings of human driving demonstrations. From the task’s perspective, real-time interaction characteristics of driving tasks make data-driven model deployment incapable and error-prone. Execution of misleading

or erratic driving behavior generated by AVs is inclined toward traffic accidents or even fatalities, which threatens the traffic safety situation. Moreover, better model performance does not equal higher reproducibility or interpretability because the relationships discovered from driving data do not necessarily indicate real-world causation and are even invalid. Hence, data-driven models are incremental to current practices because their inherent characteristics remain unchanged, i.e., fitting data correlations from extensive demonstrations without explicit physical meaning.

Consequently, there is a groundswell for designing human-like AVs and providing human-level driving thinking ability. From the viewpoint of behavioral sciences, informal driving rules and implicit driving interactions are considered human prerogatives. Thus, AVs should hold a “common sense” of human-level driving and interpret driving behaviors or interactions in similar pipelines to human drivers to become human like. More importantly, human-like AVs are supposed to acquire human thinking properties, such as top-down reasoning driven by cognition expectations or a sense of causality, to understand what to do in novel situations, which has great potential to break current research bottlenecks.

## NEW ADVANCES IN SOCIAL NEUROSCIENCE

### Interbrain synchrony

In social neuroscience, studies have started exploring interbrain (or brain-to-brain, multi-brain) synchrony among participants engaged in the same activity. Specifically, experiments revealed that group behaviors and interactions could raise coordinated bidirectional interbrain correlation patterns in animal species like bats, mice, and monkeys.<sup>1</sup> This theory was generalized to collective animal social neuroscience and has become a promising research direction toward social intelligence.

Recently, similar interbrain synchrony has also been reported in humans. For example, brainwave synchrony exists among high school seniors in real-world dynamic group interactions in a classroom.<sup>2</sup> The results suggest that interbrain synchrony may act as a neural marker for quantifying human dynamic social interactions. Later, studies on the neural mechanisms of human interbrain interactions in music listening, gesture imitation, mutual gazing, cooperative decision-making tasks, etc. have demonstrated similar positive outcomes.<sup>3</sup>

These studies shed new light on modeling interactions of human minds or studying the “social brain.” By understanding how individual and cooperative brains work from neurological views, interbrain synchrony can be modeled quantitatively by considering individual behaviors, shared cognitive states, and social contexts.

### Neuromorphic computing

Another critical research trend is developing neuromorphic or brain-inspired computing models to utilize human cerebral neurological principles or mechanisms and build biologically plausible computational frameworks. This research is envisioned to provide a novel brain-like computing technology that can learn and adapt like humans. Unlike conventional computing systems that rely on the Turing machine or von Neumann architecture, neuromorphic computing models perform computation by imitating the connections and charging/firing process of neurons and synapse potential (neural spikes). Comprehensive surveys on neuromorphic computing models and brain-inspired system hierarchy can be found in a previous study.<sup>4</sup>

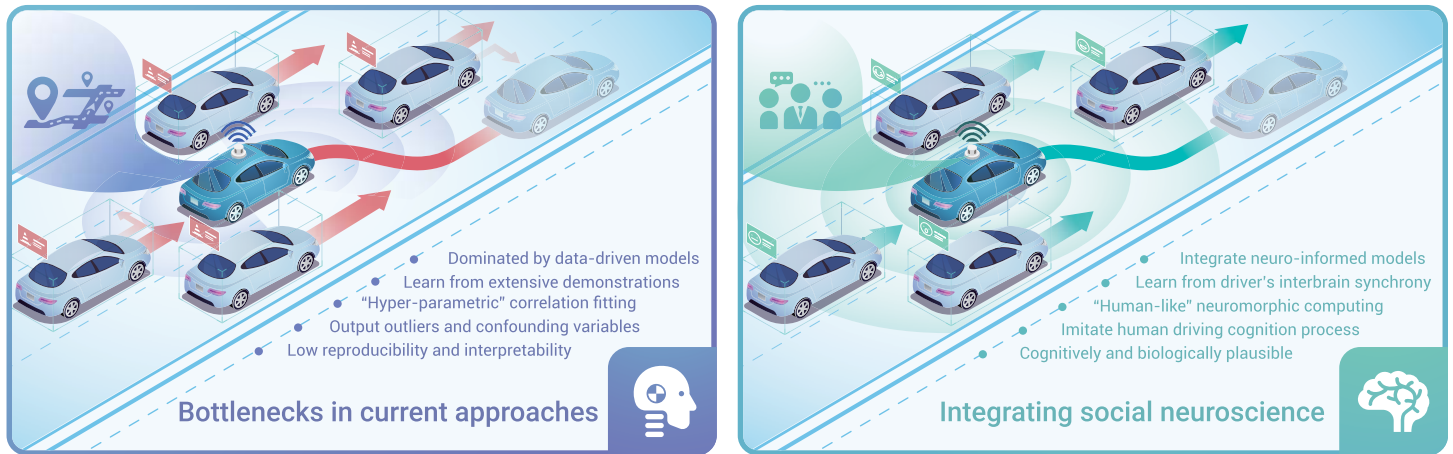


Figure 1. Integrating social neuroscience to break current driving understanding bottlenecks

Moreover, because humans can make quick and appropriate decisions based on incomplete information with top-down reasoning mechanisms (natural intelligence [NI]), its computational imitator, neuromorphic computing, is believed to outperform current artificial intelligence (AI) models under complex, partially informed, real-time decision-making scenarios. Experiments have also revealed that neuromorphic computing models are competent in a spectrum of human-involved socially interactive tasks, such as human-robot cooperation and interactions.<sup>5</sup> Consequently, this can lay a solid computational basis and act as a powerful tool for implementing human neural responding mechanisms during cooperative driving tasks, which can address human driving interactions in a human-like, biologically plausible manner.

#### TOWARD COLLECTIVE NEURO-INFORMED DRIVING UNDERSTANDING

With technological and analytical innovations, applying social neuroscience to engineering applications has become feasible, such as integrating interbrain synchrony and social neuromorphic computing research into modeling human driving interactions. Because mutual driving understanding and interaction require human drivers to cooperate in the driving task, they are assumed to share synchronized interbrain activations like other social cooperative tasks. Following this assumption, we can establish the interbrain neural response similarity as an indicator to represent the behavioral cooperativeness of drivers, which addresses the interpretability issue of data-driven models using collective neural information. Moreover, the social neuromorphic computing models can be integrated into driving understanding computation and behavior generation modules and replace the pure data-driven ones, which can yield better behavior explicability and reproducibility.

As shown in Figure 1, this collective neuro-informed driving understanding research has the following key merits. First, given interbrain synchrony and neuromorphic computation, the proposed research is biologically plausible to overcome the interpretability and reproducibility bottlenecks of current studies. Second, the proposed research reasons and computes human driving behaviors at an evidence-based neurological level (neuro-informed) rather than blindly learning or mimicking the erratic nature of human behaviors. Last, the integrated human-like AVs are envisioned to behave in a manner that is understandable and expected by other humans, which tackles mutual driving understanding barriers and facilitate future vehicle coexistence.

Unlike the straightforward proposal, this social neuroscience integration is an interdisciplinary research frontier requiring deep intersectoral collaborations from social neuroscientists, behavioral scientists, computer scientists, and vehicle engineers. Key factors and research paradigms are still under investigation. Extensive

emerging efforts should be made to advance the theoretical basis and implementation paradigms to this research.

#### CONCLUSION

In this Editorial, we discussed the possibility and merits of integrating social neuroscience into human-machine mutual behavioral understanding for AVs. There will be much work to achieve human-like autonomous driving because both research directions are still in their infancy. The research is worthy of investigation for discovering the underlying neurological nature of human driving behaviors. As long as human-like AVs and human drivers utilize similar neurological pipelines to generate decisions from similar knowledge, they are supposed to reach mutual understanding and interact intelligently.

To illuminate a bigger picture, we use this research example to advocate two future research directions for general engineering disciplines. The first is to bridge the technology gap between AI and NI by exploring and imitating human neural cognitive processes, and the second is to address human-robot interaction issues by introducing collective social intelligence. This would promote AV research to a human-like level and benefit other industries empowered by AI, machines, or robots.

#### REFERENCES

1. Sliwa, J. (2021). Toward collective animal neuroscience. *Science* **374**, 397–398.
2. Dikker, S., Wan, L., Davidesco, I., et al. (2017). Brain-to-brain synchrony tracks real-world dynamic group interactions in the classroom. *Curr. Biol.* **27**, 1375–1380.
3. Czeszumski, A., Eustergerling, S., Lang, A., et al. (2020). Hyperscanning: a valid method to study neural inter-brain underpinnings of social interaction. *Front. Hum. Neurosci.* **14**, 39.
4. Zhang, Y., Qu, P., Ji, Y., et al. (2020). A system hierarchy for brain-inspired computing. *Nature* **586**, 378–384.
5. Aitsam, M., Davies, S., and Di Nuovo, A. (2022). Neuromorphic computing for interactive robotics: a systematic review. *IEEE Access* **10**, 122261–122279.

#### ACKNOWLEDGMENTS

The work is supported by the National Natural Science Foundation of China (72288101 and 72171210), the China Postdoctoral Science Foundation (2021M702819), the Zhejiang Provincial Natural Science Foundation of China (LZ23E080002), the National Key Research and Development Program of China (2020AAA0107400), and the Smart Urban Future (SURF) Laboratory, Zhejiang Province, China.

#### DECLARATION OF INTERESTS

The authors declare no competing interests.