The Moment That The Driver Takes Over: Examining Trust in Full Self-Driving in A Naturalistic and Sequential Approach

Le Song
Department of SES
i3 Lab (CNRS)
Telecom Paris
Institut Polytechnique de Paris
Palaiseau, France
le.song@telecom-paris.fr

Zhegong Shangguan
Autonomous Systems and Robotics
Lab/U2IS
ENSTA Paris
Institut Polytechnique de Paris
Palaiseau, France
zhegong.shangguan@ensta-paris.fr

Abstract
In this paper, we have documented the challenges that drivers with autopilots experience on real-world roads by focusing on the practices of humans taking over. Using a conversation analytic approach, we analyze data of full self-driving cars selected from third-party YouTube videos. We have shown how drivers treat the car's moment-by-moment motion as actions that are projectable for potentially relevant risky outcomes and how they take over the full self-driving system in situ and in vivo, with continuous situated monitoring. We have demonstrated four typical situations in which drivers take over in the unfolding course of driving action: going too close to the front car, inappropriate speed in the local context, wrong recognition of lanes, and pedestrian priority. We argue that the achievement of human takeovers is inextricably connected to the situated organization and accountability of the course of action.

Keywords: Trust, full self-driving, human assisting, multimodal conversation analysis
1 Introduction

1.1 Human and self-driving car interaction from the conversation analysis perspective

Autonomous vehicles (AVs) are already commonly seen on the roads. In the foreseeable future, smart cars with L2 and above functionalities will become mainstream. Currently, the majority of technical challenges lie in artificial intelligence (AI) algorithms, path planning, etc. However, the interaction between drivers and automated systems, especially the conflicts encountered among human-machine interactions (HMI) in natural driving, is an area that desires further research.

The interaction between humans and self-driving vehicles presents a multifaceted challenge, as underscored by recent research. Traditional nonverbal cues utilized by human drivers to convey awareness and intent to pedestrians are absent in autonomous vehicles, necessitating the development of alternative interfaces. Mahadevan et al. (2018) emphasize the importance of multimodal interfaces beyond vehicle movement, which can aid pedestrians in understanding the vehicles' intentions, particularly in crosswalk scenarios. Drawing from cross-cultural analyses of human traffic interaction, Brown et al. (2022) identify fundamental movement elements crucial for safe maneuvering on roads, such as gaps, speed, position, indicating, and stopping. These insights suggest opportunities for designing vehicle motion to enhance understandability within traffic contexts, thus improving overall safety and efficiency.

However, autonomous vehicles still struggle with the complex social dynamics of traffic, as evidenced by Brown et al. (2023), who found shortcomings in their ability to navigate yielding situations. This highlights the challenges in designing autonomous vehicles to effectively communicate intentions and respond appropriately to other road users' actions. As a higher manifestation of intelligent behavior, social behavior is still difficult for current AVs to handle these social interactions.

Moreover, the challenges drivers face when using GPS navigation systems, as Brown & Laurier (2012) investigated, underscore the need for technology to support "instructed action" by fostering a better understanding of how drivers interact with navigation systems. This insight is relevant for designing interfaces that facilitate seamless collaboration between human operators and automated systems. Similarly, the analysis of public videos of assisted and autonomous driving by Brown and Laurier (2017) emphasizes the importance of transparency in vehicle actions to enhance mutual understanding among drivers and improve overall road safety. This highlights the need for vehicle systems to communicate their intentions and actions effectively to human drivers and other road users.

Overall, these studies underscore the complexities inherent in human-vehicle interaction (HVI) and suggest avenues for designing interfaces and motion behaviors that facilitate clear communication of vehicle intentions while ensuring safety and efficiency on the roads. Future research should continue to explore innovative solutions to address these challenges and facilitate the integration of autonomous vehicles into existing traffic ecosystems.
1.2 Trust in human-machine interaction study

Trust plays a pivotal role in shaping human interaction with machines (Harper & Odom, 2014; Ivarsson & Lindwall, 2023), particularly in the context of human-robot interaction (HRI) and human-agent interaction (HAI). Plurkowski et al. (2011) highlight the significance of interactional repair mechanisms in human-robot interactions, drawing insights from card-game activities to inform the design of autonomous social robotic systems. Their study employs a conversation analytic (CA) approach to examine how humans identify and manage interactional trouble within activities, offering implications for future autonomous robot design. Rieger et al. (2023) look into the dynamics of trust formation in human-agent interactions, comparing trust attitudes towards automated systems and human experts. Their findings suggest that trust attitude and perceived reliability are higher for human experts than for AI, emphasizing the importance of considering agent expertise in HAI scenarios. This underscores the existence of an imperfect automation schema, indicating potential challenges when introducing novel AI support agents. González-Martínez & Mlynář (2019) introduce the concept of "practical trust," emphasizing the continuous practical work involved in trust formation within interpersonal interactions. They advocate for empirical research to identify the specific practices that constitute trust as an interactional phenomenon, highlighting its strong link to the organization of action and accountability. Watson (2014, 2009) revisits Garfinkel's notion of trust (Garfinkel, 1963; Turowetz & Rawls, 2020), emphasizing trust as a background condition for mutually intelligible action. Through an ethnomethodological lens, Watson (2014) underscores the importance of understanding trust within the context of constitutive practices and everyday sense-making activities. Eisenmann et al. (2023) examine Garfinkel's research on human-machine interaction, particularly with ELIZA and LYRIC programs from 1967 to 1969. They argue that successful human-machine interaction relies on exploiting human sense-making practices rather than solely on machine characteristics. This historical perspective has implications for contemporary AI design, suggesting a need to integrate human social practices into computational systems. Ivarsson and Lindwall (2023) examine the issue of trust and suspicion in conversational agents and human interaction. By departing from ethnomethodology and conversation analysis, they illustrate how parties in a conversation understand their interactional partners. When suspicion is roused, shared understanding is disrupted. They claim that the proliferation of conversational systems, fueled by artificial intelligence, may have unintended consequences, including impacts on human–human interactions.

In summary, past related works highlight the multifaceted nature of trust in human-machine interaction, emphasizing its role in shaping intersubjectivity and interobjectivity (Latour, 1996), informing design considerations, and enacting trust as a social reality (Lewis & Weigert, 1985). Ethnomethodological approaches offer valuable insights into the practical work involved in trust formation, while historical perspectives provide context for understanding the evolving nature of human-machine interactions.
1.3 The ‘Zero Takeovers’ Challenge in full self-driving mode

![Image of a car interior with a driver looking at the dashboard]

Figure 1. FSD Youtubers are doing “zero takeovers” testing on self-driving cars, usually by keeping their hands off the steering wheel

Some full self-driving (FSD) testing Youtubers would like to take the so-called ‘zero takeovers’ challenge (figure 1). They try to put no hands on the steering wheel and the touchpad (Jung et al., 2020) and keep the vehicle driving by itself as long and as much as possible, even if it is conditions like night, snow, or hill roads.

The following case is an example showing how the tester tries to keep the car turned by itself at an intersection in a moment-by-moment way. Although trying to keep his hands off the steering wheel, the driver holds his two hands just above the wheel and monitors the car’s motions all the way, trying to achieve the accomplishment of “don’t hit anything.” He narrates and describes every essential action of the car and the road conditions, appearing to decide to touch and control the wheel at any time.
Figure 2: “Don’t hit anything”. Driver doing “zero takeover challenge” in FSD mode. I use a comic fashion visualization for the transcription of moment-by-moment interactions. Video from Marques Brownlee YouTube channel, Dec 15, 2022

Keeping the hands at a close distance from the wheel or even holding the hands on the wheel is a way of dealing with the issue of trust in FSD. Here is another case that shows the driver always holding his hands on the steering wheel because “I didn’t trust there” in his narrative of the video. The driver claims the car does not manage so well in some situations, even for the Enhanced Autopilot.
The issue emerges as how drivers deal with situations in which they do not trust self-driving cars and how they take over by adopting human assisting actions (Brown & Laurier, 2017) in situ and in vivo (Cahour et al., 2012). In this paper, we aim to examine how human drivers’ taking over actions occur as situated and multimodal practices in full self-driving mode by investigating “why take over now” from the perspective of EMCA. We will check some of the typical situations in which the drive takes over.

2 Method and data

2.1 Public online FSD testing videos

Online collections of third-party videos offer a valuable source of information on human interactions and technology use in various settings, especially with the advent of new technologies. YouTube is currently the most extensive repository of third-party videos in the world. However, it is essential to exercise caution when searching for, organizing, and reusing videos on the platform. Raw data from online self-driving car testing videos adhere to the naturally occurring principle of conversation analytic video analysis (Deppermann et al., 2018). For this study, we collected 12 hours of video from YouTube. In our data corpus, there are altogether 50 clips that come from 24 YouTubers.
2.2 Multimodal conversation analysis

We employ multimodal conversation analysis (Hazel et al., 2014; Schegloff, 2007) as the research method. Our analysis aims to identify, describe, and explain the orderly and recurrent practices adopted by participants during the driving activities. We closely examine the multimodal resources that emerge within the technologized contextual configurations, including gestures, gaze, body postures, and the embodied manipulation of all devices.

The next-turn proof procedure by Sacks, Schegloff, and Jefferson (1974) is an essential guide for the conversation analytic approach. To gain insights into the question "Why that now" (Schegloff & Sacks, 1973: 299), analysts can examine how a turn is related to the previous turn and how it is responded to. In other words, people display their understanding of the preceding turn in their next turn, which can be used for analysis.

There are two main ways of transcribing the data. The first is the typical way of combining screenshots of critical actions and verbal interactions (Extracts 1, 3, and 4). The second is we use a comic style visualization (Brown et al., 2023) with the driver(s)’ narratives or dialogues superimposed on the screenshots for the transcription of the unfolding naturally-occurring interactions (Extract 2, and the “don’t hit anything” case above).

3 Analysis: Human Assisting As Situated And Multimodal Practices In Full-Self Driving

3.1 Turning the steering wheel: “feeling a little bit too close to that car there”
01 D: I think (. ) the more you're able to monetize the software
02 (1.0) the more you can (1.0) bring down the price of the cars.
03 it actually (. ) makes sense to hook them with a low-priced
04 car#1.1( .) so that you can then (. ) hook them onto#1.2( .)
05 so#1.3ftware services#1.4.

(1.1 driving using FSD, with
hands off the wheel)

(1.2 starting taking over)(1.3 turning the wheel)

(1.4 finishing the taking over)

06 feeling a little bit too close to that car there#1.5,( .) .h so
07 that's why I took over. (2.5) B/ut you can see it did pretty
08 well (. ) navigated all the way here,
In this case, the Tesla car in FSD mode reaches almost the end of the trip under the Golden Gate Bridge in San Francisco when the driver talks about self-driving cars’ business modes. In line 4, he starts taking over (image 1.2). He turns the wheel (image 1.3), making the car head leftwards. In his later narratives, he accounts for his action (lines 6-7), saying that he feels too close to the car ahead and takes over. He also produces a positive assessment of the FSD mode that “but you can see it did pretty well (.) navigated all the way here”.

In this case, the driver treats the car’s move/action as going too close to a car/almost hitting the car. He responds by switching the car’s direction before the assumed “hitting.” This case reflects a conflict between the machine’s decision and the human driver’s decision/judgment (Park et al., 2020).

3.2 Giving a little accelerator press: “Very slowly and cautiously. Maybe a little too cautiously”

**Extract 2**
(from the “Whole Mars Catalog” YouTube channel)
The driver is in an FSD Tesla, and one of his friends sits beside him as a front passenger. The car is going through a construction zone at the moment. The car moves very slowly. The driver treats this action as being cautious and “maybe a little too cautiously.” The car even finally stopped. Confusion about intent, particularly when driving at speed, could easily cascade into collisions or other problems on the road (such as holding up traffic) (Brown et al., 2023). It is at this conjuncture that the driver narrates and gives it a little accelerator press.

3.3 Changing lanes: “It just turned down the oncoming lane”

**Extract 3**
(From “Out of Spec Dave” YouTube channel, [https://www.youtube.com/watch?v=XDxjB6bFLRg](https://www.youtube.com/watch?v=XDxjB6bFLRg))
(D means ‘driver’; C means ‘co-pilot’ as front passenger)
D: Let's put it on the FSD#3.1 beta now, here in the middle of the intersection, now it's back [up-]

C: [Whoa#3.2]

D: Oh#3.3: okay

D: .hh I'm covering the [brake,

C: [There was a red Ver[sa,

D: ]I'm covering the brake

C: it knew it was a red Ver[sa-

D: [It's like-

C: your most hated car.

D: but now there's no room GO/

C: Go

D: hahahaha

C: There we go.

D: Whoa, we're on the#3.5-[oncoming-

C: [OH:: this is not good.
In this case, the self-driving Tesla makes a mistake after turning left and goes to the oncoming lane. The drivers observe this and make the lane change (Laurier et al., 2012). Trust in self-driving cars as an observable phenomenon of order requires continuous local work even in the most commonplace situations, not to say on complicated roads. From the extract of in-car and out-car interaction, we examine the gestalt of projectable aspects of road trips, car motions, and driver talks that come to be combined with the task at hand.

3.4 Yielding for a pedestrian: “It doesn’t stop so you have to do that”

Extract 4
The last case deals with and relates to the driving rules, that is, to yield proactively for pedestrians. Since it is still hard for a self-driving car to judge whether a pedestrian is just standing by the roadside or going to pass the road, it is not easy for it to decide whether it should stop for the pedestrian (Brown et al., 2023). Therefore, a human driver needs to take over at such moments when necessary. This extract is a typical case illustrating this. The driver is driving a Tesla in FSD mode in Switzerland, and the car does not stop for a wheelchair pedestrian, and the driver has to take control of that. The two actions pair together and produce a joint activity—yielding and going on the social road (Brown & Laurier, 2017).
4 Implications

Based on the four cases we have discussed, we propose that implementing practical improvements in self-driving cars, particularly focusing on maintaining a safe following distance, adjusting speed dynamically, ensuring accurate adherence to road rules, and prioritizing pedestrian safety, holds significant implications for enhancing road safety and efficiency.

**Safe Following Distance**: Incorporating advanced sensors and algorithms to ensure self-driving cars maintain safe distances from other vehicles is crucial. Educational initiatives should emphasize the importance of safe following distances for AV operators.

**Adaptive Speed Control**: AV technology should dynamically adjust speed based on real-time traffic conditions and environmental factors. Training for AV operators should emphasize compliance with local speed regulations.

**Enhanced Road Rule Adherence**: Improving the accuracy of AVs' understanding and interpretation of road rules is essential. Developers should refine algorithms to recognize and respond to road signs and signals accurately.

**Pedestrian Priority**: Programming self-driving cars to prioritize pedestrian safety through advanced detection technologies and predictive algorithms is vital. Public awareness campaigns should promote mutual respect between AV operators and pedestrians.

Integrating these practical implications into the design and operation of self-driving cars can significantly improve road safety, traffic efficiency, and overall transportation experience.

5 Discussion and conclusion: the road towards “normal natural driving”

The journey towards achieving "normal natural driving" with self-driving cars traverses through intricate pathways of trust, the symbiotic collaboration between human operators, and sophisticated technological systems.

**Road Towards “Normal Natural Driving”**: As self-driving technology advances, the aspiration to emulate the natural and intuitive driving behaviors exhibited by human drivers becomes increasingly palpable. The concept of "normal natural driving" embodies the seamless integration of autonomous vehicles into the existing fabric of road traffic, where interactions mirror those of human-operated vehicles.

**Hierarchy of Trust**: Central to the successful integration of self-driving cars into the transportation ecosystem is the establishment of a hierarchy of trust between humans and autonomous systems. This hierarchical structure delineates the roles, responsibilities, and decision-making authority of human operators and automated systems.
Combining Brain with the System: The symbiotic fusion of human cognition with artificial intelligence represents a paradigm shift in the interaction between humans and autonomous systems. By harnessing the complementary strengths of human intuition, adaptability, and moral reasoning alongside the computational prowess and efficiency of AI algorithms, we can cultivate a synergistic relationship that transcends the capabilities of either entity in isolation.

In conclusion, in this paper, we have documented the challenges drivers with autopilots experience on real-world roads by focusing on the practices of humans taking over. Using videos of full self-driving cars selected from third-party YouTube videos, we have shown how drivers treat the self-driving car’s moment-by-moment motion as projectable for potentially relevant risky outcomes and take over the full self-driving system by assisting action in situ and in vivo, with continuous situated monitoring. We have demonstrated four typical situations in which drivers take over in the unfolding course of driving action: going too close to the front car, inappropriate speed in the local context, wrong recognition of lanes, and pedestrian priority.

APPENDIX. TRANSCRIPT CONVENTIONS
This paper’s transcription conventions mainly adopt Gail Jefferson's (2004) conventions for talk and Lorenza Mondada's (2018) conventions for embodied actions. Courier New font is used for the transcripts. The main convention symbols involved in this article are as follows:

, means intonation continuation;
. means downward intonation;
\ and / represent the falling and rising intonation, respectively;
( . ) means inconspicuous short pauses of the driver’s talk (usually less than 0.2 seconds);
(2.0) means the silence time of the driver’s talk when more than 0.2 seconds;
*--->
---->* asterisk marks the moment and the phase (when the action continues longer than just a moment) of the action;
# indicates where the screenshot appears;
= indicates the connection of two turns;
(( ))) indicates non-verbal behavior.

References


