The Halting problem: Video analysis of self-driving cars in traffic

Barry Brown
barry@di.ku.dk
University of Copenhagen
Copenhagen, Denmark
Stockholm University
Stockholm, Sweden

Mathias Broth
mathias.broth@liu.se
Institutionen för kultur och samhälle (IKOS), Linköpings universitet
Linköping, Sweden

Erik Vinkhuyzen
tokyovink@gmail.com
Work, Interaction and Technology Research Group, King's College London
London, United Kingdom

ABSTRACT

Using publicly uploaded videos of the Waymo and Tesla FSD self-driving cars, this paper documents how self-driving vehicles still struggle with some basics of road interaction. To drive safely self-driving cars need to interact in traffic with other road users. Yet traffic is a complex, long established social domain. We focus on one core element of road interaction: when road users yield for each other. Yielding – slowing down for others in traffic – involves communication between different road users to decide who will 'go' and who will 'yield'. Videos of the Waymo and Tesla FSD self-driving cars show how these systems fail to both yield for others, as well as failing to go when yielded to. In discussion, we explore how these 'problems' illustrate both the complexity of designing for road interaction, but also how the space of physical machine/human social interactions more broadly can be designed for.

ACM Reference Format:
Barry Brown, Mathias Broth, and Erik Vinkhuyzen. 2023. The Halting problem: Video analysis of self-driving cars in traffic. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23–28, 2023, Hamburg, Germany.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3544548.3581045

1 INTRODUCTION

At a huge cost - and massive engineering effort - companies such as Waymo and Cruise have launched robot-taxi services that drive passengers around areas of the US [22] and self-driving systems – such as Tesla's FSD ("Full self-driving") – have been rolled out to volunteer drivers in the US and Canada. While there are many media reports of the success, or otherwise [56, 57], of these systems, their actual performance is a closely guarded commercial secret. Self-driving cars introduce a host of HCI issues, such as how drivers might control or monitor self-driving systems [6], the challenge of designing the handover between autonomous and human control [74], attitudes towards self-driving vehicles [27] and ways of designing an external interface for road interaction [40]. Perhaps most tricky of all, self-driving cars must interact in traffic with other road users in a way that is understandable [19, 62]. Other road users might not know they are interacting with a self-driving car, but they still need to be able to make sense of the motion of a self-driving vehicle to avoid confusions and even collisions.

In this paper we analyse third-party recordings of self-driving cars in traffic, obtained from the YouTube video sharing website. These 'in the wild' videos, uploaded by drivers and passengers, offer a distinctive perspective on how autonomous driving is developing on the road and being responded to by other road users. These videos give us early access to the road as a human/machine interaction space - something of a new genre of interaction for HCI. They also allow us to understand some of the challenges that self-driving cars will face in their interactions with other road users. For this paper we define and focus on a very basic interaction that is common in traffic: yielding for other road users. Yielding – achieved by (for example) slowing down – involves communication between different road users to decide who will 'go' and who will 'yield'. For analytic purposes, yielding is a simple, almost fundamental, traffic interaction, providing us a relatively straightforward unit through which we can study how self-driving cars can interact with other road users.

We analysed a corpus of over 16 hours of driving, developing a closer analysis of a small set of examples, which we document here. We present four incidents in particular where self-driving cars 'yield' while in traffic with other human drivers and pedestrians. This lets us explore how even a simple operation like yielding contains a surprising amount of subtlety. In discussion, we explore the 'problems' that self-driving cars have with yielding, and the importance of social interaction for safe driving. We will discuss something of what is needed for smooth physical interactions in busy human occupied spaces more broadly. Stopping can be done 'in the wrong way', and just as seriously, others' stops can be misunderstood (or not really understood at all). This offers lessons for social robotics, as well as research on self-driving cars, and how to improve the design of physical social interaction.

2 BACKGROUND

2.1 Autonomous vehicles

In recent years, commercial and academic interest in autonomous driving has grown massively, drawing on advances in machine learning and computer vision, along with large industrial investments into autonomous vehicle development [2, 21]. Recent work has engaged with problems such as challenges in image recognition [73], the role of machine learning in understanding road terrain [72], as well as discussions of the socio-economic impacts and challenges these systems provoke [51]. It is important to note that while there has been impressive progress [48], significant challenges remain on basic tasks–such as using vision to locate a car [9]. Chris Urmson,

CHI '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9421-5/23/04…$15.00
https://doi.org/10.1145/3544548.3581045

one of the founders of the Google Self-Driving project, recently acknowledged that self-driving cars will likely appear gradually "over the next 30 to 50 years" [21]. It is likely then that the development of these vehicles will remain a long-term technical challenge [52].

Autonomous vehicles also present a range of human-centred challenges, and this has been a major focus of work in the automotive UI domain. While research initially focused on the inside of autonomous vehicles, for example interface issues such as how to manage handovers between the human driver and autonomous car [75], more recently interactions between autonomous vehicles and other road users has come into focus. In particular, a series of experiments have explored external human machine interfaces (eHMIs) - typically using lights and sounds to communicate an autonomous vehicle's intent to other road users [40, 46, 49]. For example, Mahadevan *et al* explored a range of different interfaces to communicate vehicle intent [49]; other researchers have explored how specific vehicles, such as trucks, can signal that they are, for example, making a delivery [11]. Merat [54] looked at how pedestrians make sense of cars at junctions, and how systems could assess the assertiveness of different pedestrians. Indeed, pedestrians generally have been a challenging topic as systems struggle to understand complex actions (such as children getting out of a car [29]). Recently at CHI, Fass, Kao and Baumann [25] documented using eHMI for cars communicating specifically with pedestrians. Yielding has also been a recent topic for eHMI work [10, 18], such as Kong et al's work documenting the different patterns in 'failed' and 'successful' yield interactions [41]. More broadly, Markkula *et al* conceptual framing [50] provides an overview of the problem of managing space and co-presence amongst vehicles.

While eHMI interfaces are a form of explicit communication, communication through movement is also hugely important. This has been characterized as "implicit interaction" – how systems communicate "implicitly" through their movement in space [38]. This is particularly a challenge in designing driver-assisted and self-driving cars, in that these vehicles must move around, and interact with other road users at speed. As Risto et al summarise: "[an] understanding of the effect and importance of movement gestures in day-to-day traffic interactions is needed for developers of autonomous vehicles" [66] (see also [1, 62, 63]). This work connects to work within human-robot interaction concerning how robots can yield to humans [35] (such as through 'backing off' [64]).

2.2 Interactional studies of driving

Recent research [4, 14] has documented the 'sociality of traffic' – the taken for granted order that makes public space effective and safe. This work has shown how even momentary interactions between unacquainted road users at speed are social. As self-driving cars enter these situations, there is the potential for disruption, conflict and danger through miscommunication. Systems will need to both understand the movements of others, but also understand how their own movements will be seen by others [4, 24, 38, 39].

Within sociology, interest in communication on the road has focused on interaction within the car, such as talk between drivers and passengers where the journey provided a context and resource [17, 45]. A series of recent articles have turned their attention more on driving, such as how decision-making about routes is accomplished [30, 44]. This led to increased interest in driving itself as a skill, with a number of studies examining driving lessons and driver training [4, 13, 15, 55, 59] . This work examines how instructors often formulate explicitly how student drivers should act in order to communicate with others, while also explaining the communicative actions of others in ordinary road situations. From these training studies a series of elements are marked out as central in the recognition of what drivers and other road users are doing. Firstly, the semiotics of the road environment itself with its road markings, signs, lights, engineered features (e.g. roundabouts, bridges etc.). Secondly, the ongoing display of current actions, intended next actions and projects. Thirdly, action-response sequences that allow for requests, offers, refusals, mistakes and reprimands. Particular actions might be marked by the flashing of headlights or not moving the vehicle forward and thereby opening up a gap ahead as an offer for another vehicle.

While useful, this work has only engaged in passing with the design problems of self-driving vehicles. So while there is a work on designing self-driving cars' interactions in traffic, this work has focused mostly on explicit interaction, and in turn the papers that have looked closely at implicit road communication have not explicitly considered the design of self-driving systems. Work by our extended authorship team addresses some aspects of this. Brown and Laurier's 2017 paper [6] focuses on 'co-driving' - where cars take over some functions of driving but rely upon monitoring work done by their human drivers. Similar to this paper, that work draws upon publicly available videos to understand how interaction in traffic happens, although with an earlier generation of self-driving vehicles (as this data was from 2016). They conclude that the "nature of driving is reconfigured through allowing the autopilot to undertake simple tasks as a co-pilot". In more recent work, Brown, Laurier and Vinkhuyzen use both in-car and street-side recordings to develop a library of road interactions for self-driving car design [7]. They discuss five basic movements: gaps, speed, position, indicating and stopping, and how together these can be combined to make and accept offers, show urgency, make requests and display preferences.

2.3 Public online videos as a data source

Online collections of third-party video provide an interesting new source of data on interaction and the use of new technologies across an array of settings (e.g. unboxing new products, tutorials, trials etc.). YouTube now forms the world's largest repository of third party video, though care should be exercised over how its videos are searched for, organised and re-used [38]. A number of papers at CHI have drawn upon YouTube data [9,32] as a resource for understanding ordinary interactions with technology. In the social sciences, YouTube data fits well with established conventions of using secondary data and connects with recent initiatives to collect large scale qualitative data sets.

Collecting and analysing third party recordings is a rapid method for obtaining early adopters' perspective and user reports. Indeed, third parties videos can provide access to events, and a diversity of events, which planned data collection would find almost impossible to capture. Third party video collected from online sources has been used previously to study policing practices [37], inter-generational communication [32], argumentative communication [65] and mobile device use [36]. Using the YouTube videos of self-driving car drivers or passengers to study traffic does mean a loss of control compared to methods such as equipping drivers with cameras (e.g. [6]). Videos recorded from inside a self-driving vehicle also have inherent limitations, as they do not provide access to the viewpoint of other drivers or road users, and so (for example) we cannot know if other road users are aware they are encountering a self-driving vehicle. However using YouTube recordings has clear advantages in terms of scope, access, diversity and availability.

3 DATA

Drawing on the approach laid out in Brown and Laurier's paper [6], we turned to YouTube as a resource for obtaining data on self-driving cars. Most self-driving systems that are being tested have YouTube videos of varying quality and quantity - including GM's Cruise, Baidu's Alpha and Pony.AI. Yet in terms of videos materials available online two companies stand out: Waymo and Tesla.

Waymo has clearly one of the most advanced self-driving systems, and is one of the few companies to run an open access robot-taxi safety-driver free service in Chandler, a suburb of Phoenix. By now, Waymo's 'robotaxis' have covered millions of miles on public roads, at all hours of the day, without major accident, a very impressive feat. However, the environment in Chandler is perhaps one particularly suited to autonomous driving, as it is a modern American suburb designed for driving, with wide and well-marked roads, little in the way of problematic weather, and few pedestrians or bicyclists. Waymo has now expanded its testing to several other cities, but usually with safety drivers behind the wheel [23]. Having a service open to the public means that there is considerable video data online of the Waymo car in use as a robotaxi. While we have reviewed other creators videos of Waymo drives, for this paper we will rely on the videos posted by one Waymo passenger in Chandler, JJ Ricks, who has recorded and uploaded 70 or so videos from journeys with the Waymo self-driving cars. These videos feature long, unedited recordings of the Waymo robotaxi navigating Chandler, with multiple cameras, usually one pointed out the front window and one pointed at the screen that visualizes parts of the internal state of the software. This produces a collection of videos which fit well with our research purposes to analyse the performance in a wide variety of situations.

Tesla is taking an alternative approach; at the time of writing this paper they have allowed around 100,000 US and Canadian based drivers access to their "Full self-driving" (FSD) beta software. To select these drivers, Tesla scores their customers' manual driving and picks only 'safe' drivers–those who score high enough–to test the software. Tesla customers have tested the FSD software in their electric vehicles in a wide variety of environments, including rain and snow, busy downtowns, as well as country roads. Tesla captures the logs of the drives when beta tester drives use FSD – gaining access to the software explicitly involves a user enrolling in the 'beta test' and providing data to Tesla - and applies machine learning to the data improve the software.

Capitalizing on the keen public interest in self-driving cars, a number of beta drivers record their FSD drives and post them on YouTube. Many of these videos feature relatively long recordings of drives, and to evaluate the software 'scientifically' a number of uploaders repeat the same drive every time a new version of the software comes out. While these videos can be revealing, most of them are taken in a suburban US road environments and only a few test the car in more challenging city environments. The video materials can be very sophisticated. One notable creator, Chuck Cook, has recorded hundreds of cases of the FSD software attempting a particularly difficult left turn across multiple lanes of traffic, even using a drone to provide a stable 'overhead' view of the intersection that he synchronises with his in-car video and commentary. In one tweet, Elon Musk (CEO of Tesla) mentioned that the latest release of the beta software should be better at what he called "Chuck's corner" – referring to this left turn [12]. In response to Chuck's videos, Tesla had sent test drivers to that specific corner (and others nearby) to tweak the software so that the next version could better handle the challenging unprotected left turn. In this way YouTube content creators have become an integral part of the actual development of the Tesla's self-driving software. While this is a testament to the seriousness of the testing undertaken by Chuck and others, it also points out that these videos are made for a specific purpose and should not be taken to be a neutral rendering of what happens or would happen 'in the wild'. Nevertheless the quality of the video materials is often very good, and excellent data for empirical investigations of self-driving car behavior on the road.

4 METHOD

As researchers with a ethnomethodological perspective on activity [28, 34], we wanted to take an approach that documented the 'seen but un-noticed' aspects of driving interaction. As in related work in HCI these findings often have interesting implications for design [34]. Our approach in particular was informed by the longstanding tradition of work within HCI and CSCW that uses video to look closely at the moment-by-moment interaction with technology [8, 26, 33]. We also draw heavily on the long history of conversation analysis as an approach to understanding human interaction. This work has pioneered looking at interaction in terms of sequences of action, with an intense focus on small sections of data in an attempt to provide a 'deep' rather than 'broad' or summary analysis of the phenomena.

We should emphasise that we are not attempting to evaluate either system in any sort of direct way. Instead we wish to understanding what 'road interaction with self-driving-cars' is, as a human computer interaction research topic, and in particular what challenges are outstanding in designing self-driving cars. We started by building a corpus of video data of both the Waymo and Tesla self-driving tests. We downloaded all the JJ Ricks test videos, 70 videos of around 15 minutes each, with a total runtime of over 18 hours. We choose JJ Ricks because of the relatively long and unedited nature of his videos, that provide us access (mostly) to the whole ride for the rides he records, allowing us unedited access to the behaviour of the Waymo car.

Dividing the videos amongst the authors we watched each video, extracting incidents from the videos focusing on cases of road interactions between the Waymo and other people. From these we selected a collection of 40 interaction 'highlights' - looking for cases where the interaction seemed noteworthy - either because something went wrong, or an interaction was particularly smooth, or cases that seemed particularly unusual or typical.

With the Tesla FSD videos we selected five FSD YouTube uploaders (Dirty Tesla, Whole Mars Catalog, AI Drivr, Chuck Cook and Kim Paquette) who recorded and uploaded relatively un-edited drives with FSD, with particular interest in those videos with traffic interactions. From this we downloaded around three months of their most recent videos, and after further group analysis we selected a smaller selection of 12 videos for more intense analysis, with each of these drives being around 20 minutes. Again, from this we selected clips where the road interactions were smooth, problematic, or seemed unusual or typical. The fragments were analysed in group data sessions [33]. Our analysis took the form not of the application of a formal method, but a more crafted set of analysis sessions and informed inspection of clips. Each extract was thus looked at as an individual, unique incident – but also inspected for exemplifying patterns that we could extrapolate to understand road interaction more broadly. Our analytic approach builds on earlier analyses of how drivers co-ordinate and manage their driving with each other [5, 13] and on Watson's [76] studies of specialized forms of driving.

For this paper we focused on one road maneuver in particular - yielding - and went back to our broader collection to extract a smaller collection of about 20 'yield' incidents from the Tesla and Waymo data. We then re-analysed these clips drawing both on the literature on yielding in traffic, and through comparisons between the different clips and our observations of self-driving behaviour more broadly.

Ultimately we selected four clips that illustrate both how the Waymo and Tesla drive, some of the problems they encounter, as well as how they can both drive successfully for much of the time. These clips give a schematic overview of our broader dataset, but we also selected these clips to display some of the more distinctive aspects of road interaction. The clips presented below are only a small selection of all the videos those we analysed for this paper.

5 RESULTS

For much of our data both the Tesla FSD and the Waymo cars drive without any noticeable problems, navigating different junctions and traffic with other drivers. However, clearly both these systems cannot yet drive without external support (be that remote or in-car). In the Waymo dataset of around seventy rides there are three interventions where the system disengages from driving, or gets stuck in some way. In one notable case, the Waymo stops and gets stuck in the middle of a busy road and has to be retrieved by Waymo support. The Tesla FSD software, although it is developing quickly, is much less robust with a disengagement or driver take over roughly every 15 minutes or so. While the content creators for the Tesla videos are broadly supportive of Tesla's efforts - as the titles of the videos suggest ("Perfect 19 minute drive with no human input"; "Amazing Breaking for Pedestrians") at the time of writing the Tesla FSD system still has some way to go before it is anywhere near human driving performance.

During our analysis we became particularly interested in the issue of self-driving cars stopping in traffic - it seemed as if self-driving cars stopped in a different manner than human drivers. Drawing on the previous literature on road interaction we identified this as an issue with yielding, and this became the organising theme for the data we will look at here. Yielding to another road user can appear a deceptively simple element of driving. Yielding is a basic operation on the road, where one road user indicates, slows, holds back or stops, communicating to another road user that they can go. Haddington and Rauniomaa discuss how "road users communicate and negotiate their spatial copresence and mobile trajectories by designing their mobile actions so that they are recognizable to other road users as doing yielding." [31] In Haddington and Rauniomaa's account yielding can make use of mutual gaze and gestures between road users:

> "When road users are within each other's visual range they may draw on mutual gaze and gestures that are typical to face-to-face interaction, [a] driver tries to communicate that he is offering space to the second car waiting at the intersection. [...] The two vehicular units can thus be interpreted as both offering space—one yielding and one offering space as a form of traffic civility—and they negotiate which one of them is going to take up the offer and enter into the available space in front of them."[31]

Yielding is a social action – collaboratively achieved between road users who *yield* and road users who *go*. So yielding has two parts – a yield and a go. Accomplishing this pair on the road usually involves some sort of communication between the different road users – such as slowing, stopping, or alternatively accelerating, maintaining speed or changing trajectory. As we will see this simple operation can have some surprising complexity – such as multiple road users yielding or going, with different ways of communicating the 'yield' or 'go'. We can contrast yielding with halting in traffic – for example an emergency stop where the car stops to avoid a collision. A halt is not primarily communicative (although it has that role too) - the goal is to stop moving the car.

We discuss four examples of the Waymo and Tesla cars yielding. These clips build upon each other to produce an understanding of yielding as part of road interaction, and machine/human interaction on the road. In our first clip we look at a Waymo car (and passenger) interacting with pedestrians, in our second we look at a Waymo partially producing a 'yield' to another road user, in the third clip we look at a Tesla car (and driver) yield to a pedestrian, and lastly we look at a case of a Tesla failing to 'go', not taking its proper turn at a four way stop intersection.

5.1 Failing to yield: 'Sorry self-driving car'

In our first example (figure 1), a Waymo self-driving taxi approaches an intersection where it plans to turn right. Our YouTuber (JJ Ricks) is sitting in the back of the taxi, and there is no safety driver in the front, while JJ records video from the back seat. Ahead a group of three adults with two small children on toy cars are approaching the edge of a sidewalk. The family (we assume) is spread out, with two adults in front with the two children and a second adult walking a bit behind wearing a red shirt. As the car passes, the adult wearing

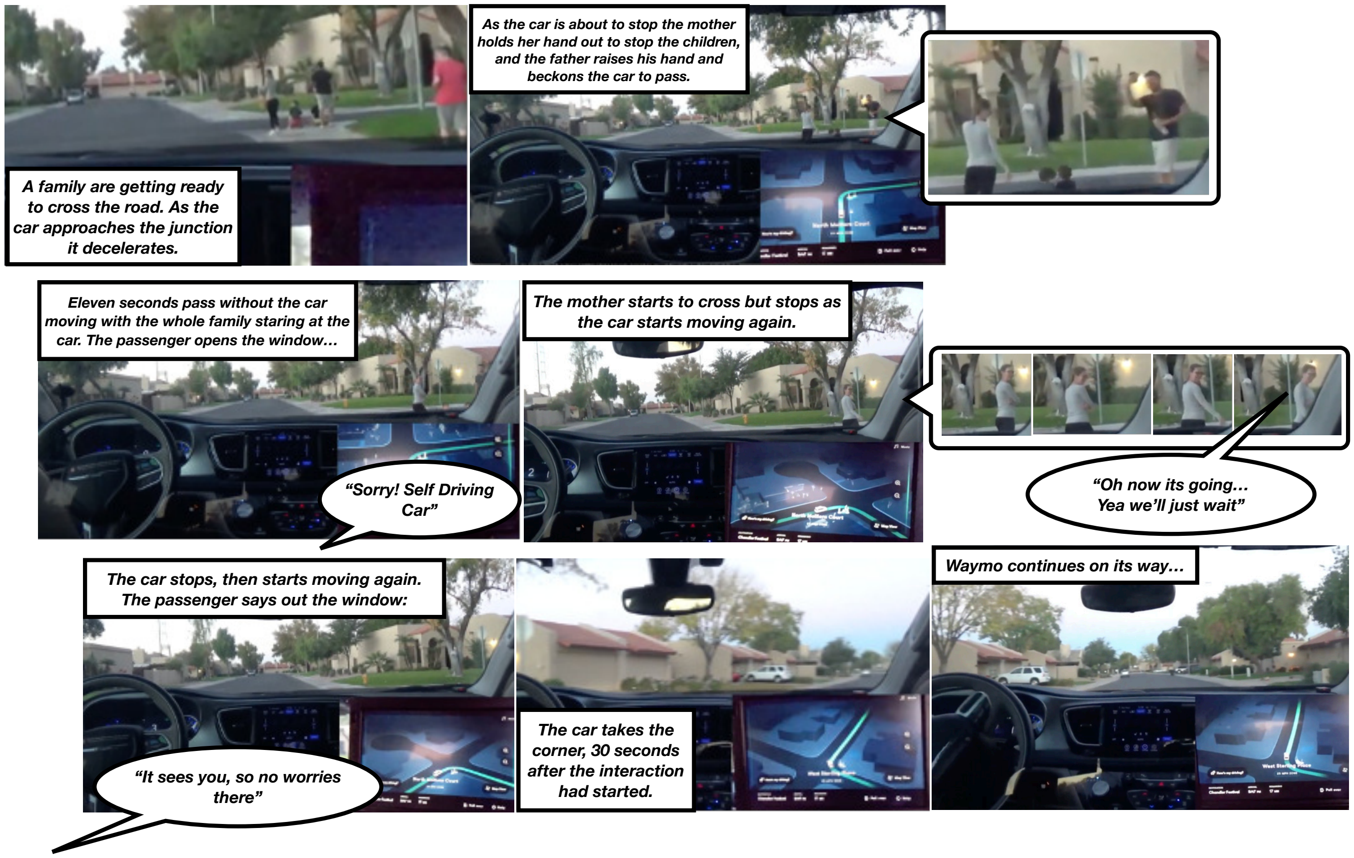

Figure 1: Waymo car and passenger interact with pedestrians. In our figures we have used a comic book visualisation for transcribing the unfolding activity [42]. Video from JJ Ricks YouTube channel, 22 October 2020

the red shirt turns around to look at the car, possibly noticing its lack of a visible driver. As the car slows down more, the ’father’ on the right with the children turns around and waves the car on, waving with his right hand in which he holds a bottle. It would be hard to find a more iconic yield gesture than that. At this point the woman stands slightly ahead of the children in the toy cars, with her hand extended down toward the children (perhaps poised to stop them should one of them go). Despite the father’s yielding gesture, the Waymo continues to slow down. The mother turns her focus from the children towards the Waymo car and also waves it on. The children now also turn to the car, and the whole family watches as the Waymo comes to a complete stop.

The family waits (the mother crosses her arms and steps back on the curb, executing a canonical waiting gesture). The Robotaxi moves forward a little, but then comes to a stop once more. At this point, around 11 seconds have passed from the father and then mother gesturing to the car, but the car is not moving, and the passenger (JJRicks) rolls down the window and says: “Sorry self-driving car”.

The mother then moves to cross and steps off the curb, but just then the Waymo starts to move also, so she abandons her crossing and steps back to the curb and says “Oh now it’s going. Yea, I’ll just wait,” withdrawing from the interaction. The car then accelerates slowly, and JJRicks again speaks out the window to the family: “It sees you, so there’s no worries about that”. The Waymo finally makes the right turn and continues on its way.

During this brief encounter between the robotaxi and the family, there are multiple moments of misalignment. When the father waves on the approaching car, the vehicle slows down instead. The mother then waves too, reinforcing the gesture, but the car comes to a complete stop. The mother crosses her arms and steps back on the curb, adopting a waiting posture. The passenger then rolls down his window and says “Sorry self-driving car”. The mother now steps into the street, but just then the car actually starts to go. She quickly retreats back to the curb, and waits until the car completes the turn.

The initial problem here is that the car does not respond to the family’s yield, implemented by stopping at the edge of the curb and then waving the Waymo on. The timing of the gesture is after the family has stopped, and may have been touched off by the father noticing that the Waymo was slowing down. While the family continues to wait at the curb, it is noteworthy that the passenger himself realizes the misalignment between his vehicle and the pedestrians, and his comment "Sorry, self-driving car" would appear to be an attempt to account for the errant driving. However, the misalignments continue. Just after his apology for the car's behavior the mother starts to cross and the car accelerates at the same time, so she quickly retreats to the safety of the curb.

Again, drawing on Haddington and Rauniomaa there are two parts of a yield – first the offer: "a mobile action (e.g., slowing down) that is recognizably offering space", followed by a second pair part that responds to the first - "the recipient's use of the offered space". These two parts produce "a fleeting and momentary orientational framework, that is, an "interactional space between road users" [58] – one that is achieved "by positioning, velocity, and various multi-modal resources, such as gestures, gaze, or car technology." [ibid] Two actions pair together and produce a joint activity – yielding and going. The car here repeatedly fails to produce the appropriate second pair part (either going or yielding) that would enable joint activity to be produced. The pedestrians wave to the car and stop – offering a space, yet the car just stops. Both the pedestrians stopping, and the wave acts as an offer – yielding to the car. But in response to this the car stops, not the appropriate second pair part. This leaves the situation in a condition of uncertainty and confusion. Should the pedestrians go? Is the car now yielding for them? At this point the car stopping could act as a different offer – it could be the 'first pair part' - the car will stop and yield to the pedestrians crossing the road. But even after this the car fails to produce a second pair part as it actually starts to slowly accelerate into the intersection just as the mother start to cross. After this the group finds themselves at an impasse – unable to work out who will cross first after two failed attempts to produce a yield. It is at this point that the passenger apologies for the car's behaviour - "Sorry self-driving car".

A related problem here in the car understanding the yield comes from the group nature of the family. The children are 'with' the adults, and so a gesture from them (and the protective positioning of the adults) can be taken to mean that all the group will wait for the car. Understanding who is yielding involves also seeing the people together as some sort of group (discussed as a 'mobile formation' in the literature [53]) - and seeing who is together involves considerable interpretation of who 'goes' together with whom [67].

This clip demonstrates how the self-driving car failed to produce either a go or yield, but the car does actually eventually get through the interaction, despite the confusion it leaves in its wake. Clearly the Waymo can drive, although it depends upon the flexibility of the pedestrians to adapt and manage to its 'witnessable incompetence'. It also exhibits how the car fails to produce this basic unit of road social interaction - and how this can cause frustrations for other road users. The pedestrians need to resort to hand gestures to signal the car on, and passenger needs to offer an account to the pedestrians of the cars odd behaviour. While this interaction is safe, accidents often are the result of an accumulation of small mistakes. It is not difficult to imagine how another road user could misinterpret the Waymo's actions, or how this kind of behavior could give rise to frustrations if self-driving cars of this sort were common on the road.

5.2 Successful yield vs programmed pattern

With this first clip introducing the concept of a yield, let us move on and look at a successful road interaction by the Waymo. In our second clip (figure 2), we join the Waymo car as it is driving along a multi-lane suburban road with the car driving in the right-most lane, past a number of entrances to local businesses. With a red traffic light ahead, the Waymo car slows with the traffic as the traffic starts to back up. However the Waymo does not stop just behind the car in front (as it would normally), but leaves a substantial gap between it and the car in front, just as there is an open junction on the right. This allows a car that is waiting at that exit to move out into traffic.

As the car stops just before the junction, the driver of the black car approaching the junction from the right is visibly looking at the Waymo car. After the Waymo car stops completely, the other driver looks forward and then drives their car out into the intersection. The car does not completely fit into the lane at first and so sits partially in the intersection, but when the traffic starts it completes its turn and the Waymo then moves narrowing the gap between it and the black car until they are both in traffic.

Unlike our first clip, this example does have the features of a successful yield. The Waymo car could have continued forward to take its place behind the front silver truck, blocking the movement of the black car into the junction. Yet the Waymo car actually slows and comes to a complete halt quite a bit before the junction (as can be seen more clearly by the LIDAR visualization that the Waymo provides inside the car).

Since cars in traffic do not usually leave large spaces between them and cars in front, stopping with such a gap acts to communicate to other road users that they are doing something specific - in this case, yielding to another car allowing it to come out into traffic. Here *where* the Waymo car stops is important in how it both physically leaves a space for the car to move out, but also how it can be seen as a deliberate action – clearly stopping before the junction, so that it has not just 'accidentally' left space for the car to come out, but is rather communicating that the Waymo car is yielding for the car at the junction to come out into traffic. As we described above in terms of the 'paired action' of yield/go, the Waymo here 'yields' and the black car then 'goes', with the yield implementing by stopping short.

One puzzle here is that while a successful yield interaction, without access to the Waymo's code it is not clear if the Waymo *has* actually been designed to yield to other road users in this situation. Indeed, we have some suspicions that the Waymo could be stopping here just because it has been programmed to leave space if it stops where there is an exit at the side of the road. In other videos of the Waymo slowing in traffic we have noticed that if there is a junction at the point where it is slowing, the car will leave an extra space between it and the car in front, so that it leaves the junction free – even if there is no car present (figure 3). It may be that the Waymo simply always stops before exits, whether there is someone to yield to or not. Again, it is not hard to imagine cases where this could be consequential. For example, if the car in the intersection had decided not to turn into traffic and to instead move when traffic has started there could be a conflict with the Waymo car who would

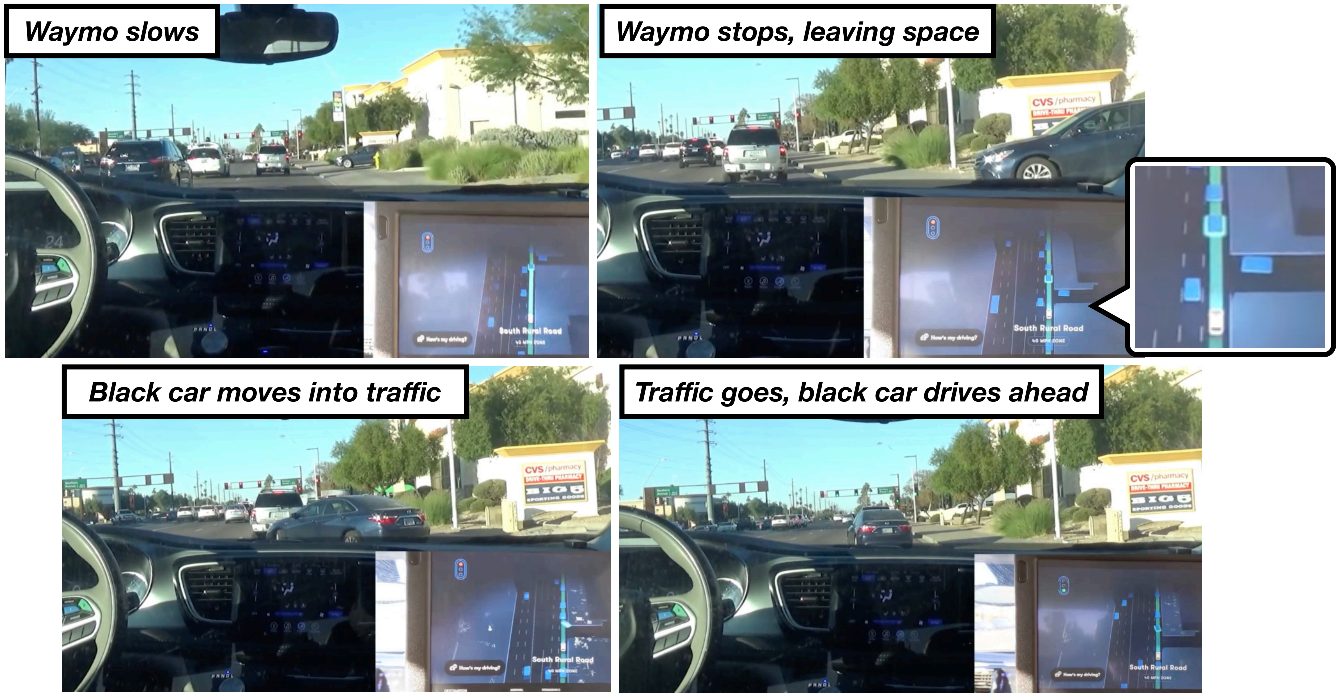

Figure 2: Waymo car yields for a car coming out of a junction. Video from JJ Ricks YouTube channel, 25th November 2020

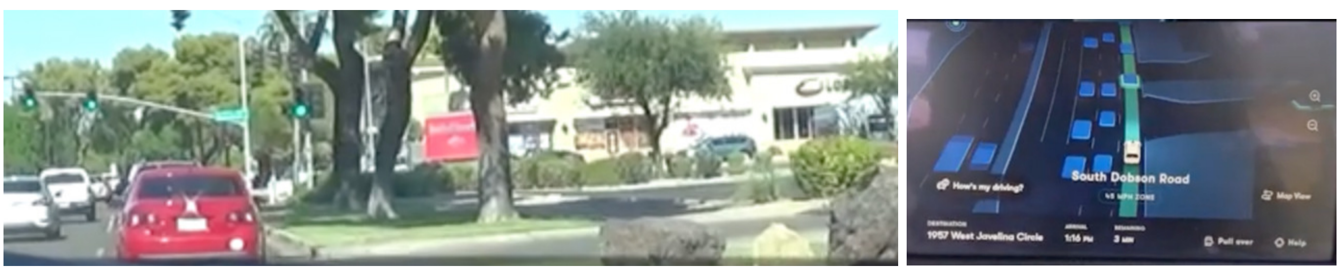

Figure 3: Waymo car is stopped in queing traffic, just before a car park exit. The car leaves a four car gap between it and the car in front. This is yielding but with nobody to yield to. Video from JJ Ricks YouTube channel, 14 October 2020

itself then also start to move - which could lead to the Waymo or other car having to stop to prevent a collision.

This raises a broader issue about the movements of self-driving cars. Cars have limited degrees of freedom, with much of driving following the simple rule 'drive if your way ahead is free'. But those movements are seen by other road users and interpreted in a social frame as if another human is driving that car. As with much orderly social human behaviour actions are read, in context, as indicative of courses of action, intent, requests, and so on. This means that if a self-driving car moves simply because its code says it is safe to do so, the car's motions will also be potentially read by others - a framing that may be completely missing to the software. This leads to a fundamental disconnect between software which is focused on driving safely, and other road users who are trying to work out 'what are you doing here'. This second clip then lets us see how there would be a difference between stopping, and properly yielding for another car, as opposed to just stopping programmatically with no understanding of how another driver might see that maneuver. In our third clip we move on to see how this can be more complicated when a driver is present.

5.3 Yielding for a pedestrian: "Got my foot over the break just in case"

We now move from the Waymo to look at our second self-driving car: the Tesla FSD system. As discussed above the Tesla FSD system is in 'beta test' status, with many thousands of drivers testing it on public roads. Unlike the Waymo, Tesla's always have a driver present who can take control of the car at any time, and so FSD driving at this stage is better described as a driver assistance feature, although the eventual aim is for the system to be a form of unassisted self-driving. This setup means that our Tesla clips offer something of a three agent setup – with other road users, the FSD

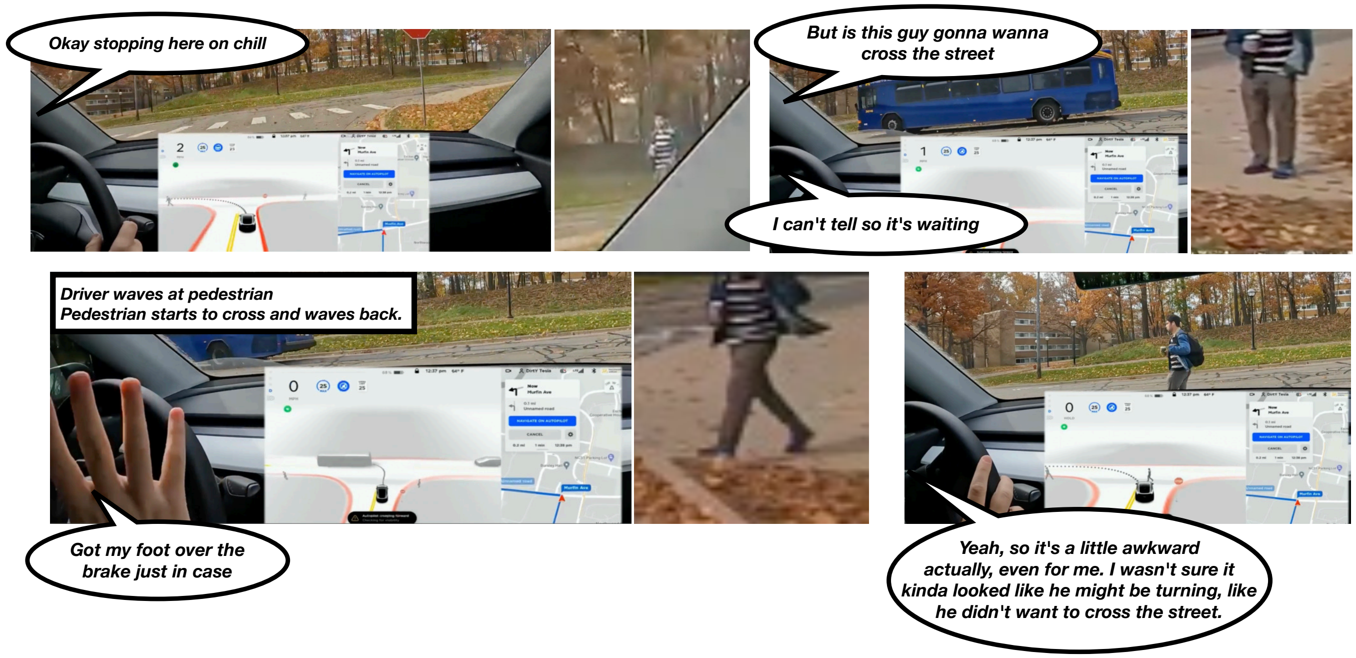

Figure 4: Tesla yields for a pedestrian. The pedestrian is visible from a second camera, which we have cropped and included at the right of each in-car frame. Video from Dirty Tesla's YouTube channel, FSD Beta 10.4, 14th November 2021

software and also the driver themselves interacting with each other. While the Tesla software has been developing quickly (and the versions that we are looking at here are already out of date) at the time of writing it struggles with many common traffic situations, revealing again how challenging self-driving is.

In figure 4 a Tesla approaches a junction at around the same time as a pedestrian (just visible at the top right corner of the first picture). The Tesla is driving itself, using its 'full self-driving' beta features. As it approaches the junction, the pedestrian also approaches the junction. As the Tesla reaches the intersection, it slows and almost come to a full stop. The pedestrian then has come closer, but is not yet at the intersection. The driver gives a running commentary for his YouTube audience, saying "Okay, stopping here on chill." The Tesla displays a message: "AutoPilot creeping forward: Checking for visibility." Then the car starts to roll forward, and the steering wheel turns counter-clockwise. The pedestrian stops, with his body turned towards the Tesla, and the driver continues narrating: "But is this guy gonna wanna cross the street? I can't tell, so it's waiting." The Tesla then stops completely, just before the crosswalk. The pedestrian lifts his right leg and starts to walk. The message on the internal display still reads : "AutoPilot creeping forward: Checking for Visibility".

The pedestrian slowly walks towards the edge of the sidewalk, when the Tesla driver raises his right hand to the pedestrian. The pedestrian returns the gesture with his left hand as he increases his pace, now looking forward. The driver says: "I gave him a wave, so he knew" and the pedestrian crosses in front of the Tesla. The driver says: "I have my foot over the brake just in case."

The crucial point in this clip is where the pedestrian approaches the junction just as the Tesla stops. As Haddington and Rauniomaa comment on a similar example: "the pedestrian halts for a moment beside the car. The pedestrian interprets the car's stopping again as an offer of space: She begins to lean her body and move forward, and she also raises her hand to a wave in order to thank the driver and to mark the encounter as about to be closed." [31] Stopping when one has potential access to a space in front of you acts as an offer – 'you go first'.

In this case, the pedestrian interacts with the Tesla but also the driver. The Tesla starts creeping forward indicating it will be going ahead of the pedestrian, but then comes to a stop just in front of the crosswalk, thus interrupting the "going" action it had just initiated. The practical question for the pedestrian is whether the Tesla has just stopped to let vehicle-cross traffic pass, or whether it also intends to wait for him? The pedestrian's tentative step towards the crosswalk can be seen as a request, and the subsequent wave by the driver a response that demonstrates that he will wait for the pedestrian. The driver's comment to the video audience: "I gave him a wave, so he knew," is a indication that for the participants themselves the precisely timed gestures establish a common understanding: the pedestrian can cross first. His additional comment that "I have my foot over the brake just in case" indicates, too, that he also realizes that the FSD software isn't part of this interaction. If it would start to move he would have to immediately override to prevent it interfering with the just-established, mutually-agreed-upon yield. Although the driver does not actually have to interfere–the FSD software waits for the pedestrian–this could hardly be considered a pure instance of an interaction between a self-driving car and a pedestrian. Clearly the driver intervenes in negotiating the travel order, and not just the software.

This clip exhibits a similar ambiguity as the previous two: when the Tesla stops - does it actually yield, or is it just stopping for other traffic? It is common for Tesla FSD to stop for pedestrians even if, as in this example, pedestrians are 'yielding' to the car. In this case the car first appears to be going, only to then stop. No matter what the software is doing, the driver engages in an interaction with another road user, and monitors not just the cars behavior, but also how the pedestrian responds to the car's behavior, and their own gestures to each other. As with out first clip there is a clear sense of responsibility towards other road users. Note however that in the first clip the apology did not resolve the impasse, whereas in this case the driver makes certain the car will let the pedestrian go.

5.4 The California Stop: Making a mess of a four way

Four-way stops are a common traffic arrangement in North America, where a crossroads is controlled not by traffic lights or a roundabout, but by requiring all vehicles to stop and yield. Four-way stops are challenging for self-driving cars because while they do have a fixed set of traffic rules, they also have local expectations and active negotiation between road users. While road users are usually required to come to a full stop at the junction, drivers will often slow down and 'roll' though the intersection rather than coming to a full stop - colloquially known as a 'California Stop' [20]. Tesla themselves faced a design challenge in that to improve the cars performance at four-way intersections they implemented a 'California stop', but had to then remove this after some controversy and intervention by the US road safety authority [3]. Yet, as we will see in the next extract, passing through an intersection is not just about stopping, it is also about *going* at the right time.

In figure 5 the Tesla FSD struggles somewhat to get through a four way stop in San Francisco. As the clip starts the Tesla is driving up a hill and as it reaches the stop sign, with one white car already passing through the interaction, and a second white car on the right briefly waiting at the entrance to the junction. At this point either the Tesla or the white car could potentially cross over the intersection. While the white car has priority ("give way to the right"), the Tesla arrived at the junction just prior to the white car. Although drivers often give way to the first car to reach the junction, if a car is hesitant in crossing the junction this can be seen as 'yielding' by other drivers, and in response, they may enter the intersection first. This happens multiple times in this clip.

The Tesla drives at 1 mph for a short time and then stops at the stop sign, giving the white car the chance to roll through the intersection in front of the Tesla. The white car then does not stop but rolls through the junction slowly (a 'California stop').

A second blue car driving just behind the white car then comes up the junction. The Tesla does not move. Because the Tesla fails to 'take its turn' – it does start to move, but only slowly - the blue car from the right speeds up and drives across the intersection. A third car from the left (another white car) is also stopped waiting to enter the junction. As the blue car leaves the junction, this white car (car three) starts to cross from the left, blocking the Tesla's way. A fourth car (a black car) from the right approaches junction and slows. As the white car leaves the junction, the junction is clear but the Tesla does not move and so the black car enters and crosses the junction. So far there have been four cars that have taken their turn before the Tesla.

Finally, the Tesla starts to move and enter the junction – but another blue car on the right enters the intersection and crosses the intersection, causing the Tesla to halt. The driver can be seen also looking out the window at the Tesla as they drive through. As the blue car leaves on the left, the Tesla tries to move again, but it *again* stops as a white car on the left starts to enter the intersection. This white car stops when it see that the Tesla is entering the junction, but the Tesla also stops. After this second pause the Tesla starts to move again, and finally the Tesla moves across the interaction and crosses over to the other side. While the Tesla does eventually manage to get its way across the intersection, it only does this by missing its 'turn' on the junction four times, and moving and then braking three times. It only manages to go across in the end by moving while another driver is actually in the intersection.

In our first clip we discussed the importance of paired 'yield/go' actions – and that confusion is produced when the pair is not produced. Clearly we can see here that the Tesla never really manages to produce an adequate yield or go action, and it is only through a sort of 'brute force' that it actually gets through the intersection. The challenge for the Tesla in getting through this intersection is that it can neither adequately produce 'first pair parts' (where it yields for another car), nor can it adequately produce a 'second pair part', where it manages to go when another car offers and yields for it. To do these actions you must wait and go (a yield) or go and continue to drive (go). But the Tesla repeatedly goes, stops, then tries to go again and stops again, leaving the other road users unsure of what the Tesla is doing.

Conversation analytic work, on which we draw for this paper [69, 70], has explored at length the sequential nature of human interaction. For many human acts (such as turns in conversation) conversation analysis has specified how the order in which things happens matters – responses come after invitations, answers come after questions and so on. This basic observation, although it can seem to border on the trivial, actually unlocks our understanding of a whole host of different activities – a way of understanding how joint social action is mundanely, routinely produced. But as recent work by Mondada [60, 61] and others [16] has shown, the assumption of a strict sequentiality is something of a simplification. Since interaction happens in real time pre-actions often take place before the main action, that is to say pre-emptive actions that project what an other partner will do before they actually do it. There are ways in which interactants 'jump the gun'. Indeed, rather than seeing interaction as sequential it may be better to see it as a continual unfolding of reflexive adjustment between partners, with sequentiality arising from these reactions. As Mondada puts it:

> "Contrary to the rather linear successivity of [...] turns at talk, where the next turn follows the previous one, here we observe an early responsive action of the recipient. It also reveals that temporal relations between a first action and a second action in response

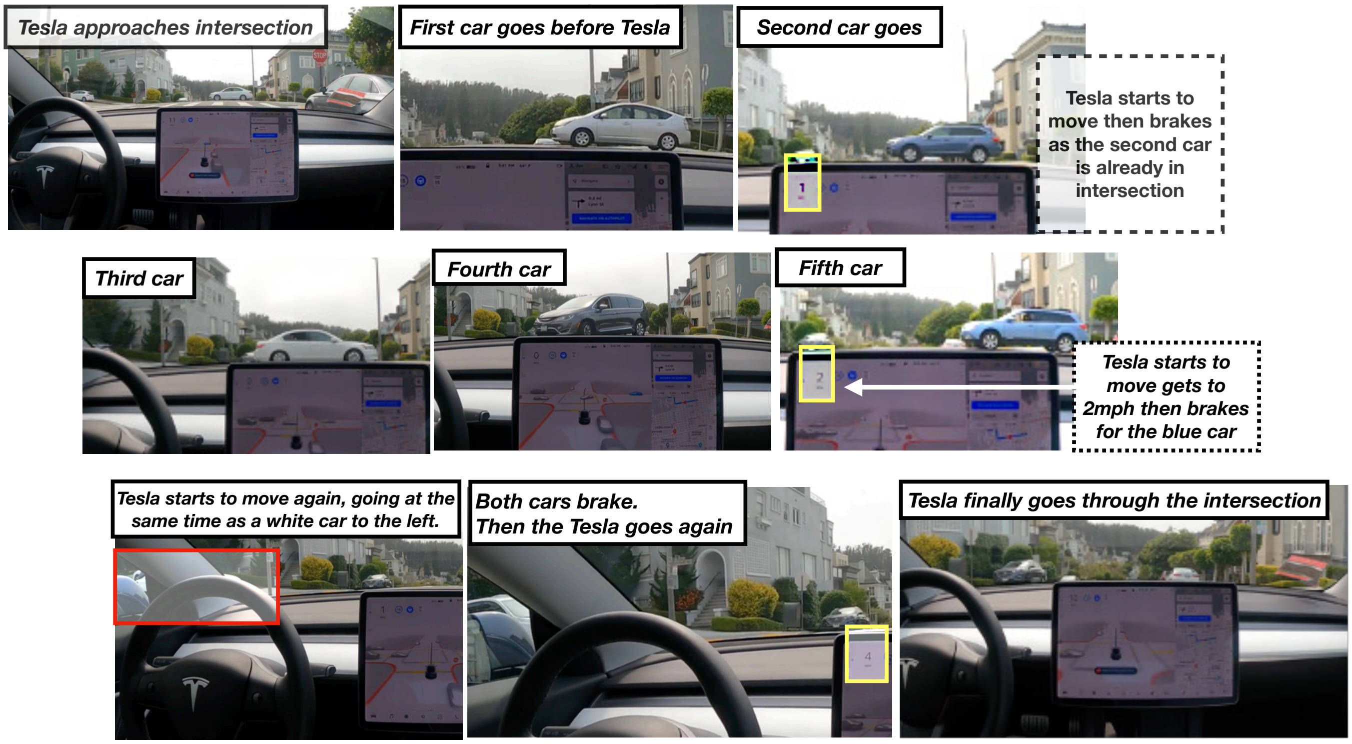

Figure 5: Tesla attempts four way stop. Five cars go before the Tesla gets through the intersection. Video from Whole Mars Catalog's YouTube channel, FSD beta version 9.2, 16 Aug 2021

> can be much tighter and complex. responsive next actions emerge as the first one is still unfolding, moment by moment, and reflexively adjusting to it." [61]

What this means, for situations such as this intersection, is that drivers' actions are not discretely bound together in terms of a yield/go but are rather continually unfolding. Road users project the future actions of other road users - not just what they are doing now but what they will do next. This introduces an important temporal character to road actions. The inability of the Tesla to produce joint action is not just because it cannot produce a second pair part (or a first) – which is bad enough on its own - but also because it cannot successfully monitor others behaviour and reflexively adjust its own to produce an action at the right time. We can see when we look at how drivers go through the intersection that they adjust their speed in concert with the Tesla's motions to try and get themselves through the intersection while still being able to yield in case the Tesla decides to go at the right time and with the right acceleration.

One point in the above interaction is revealing of this inability to reflexively adjust driving behavior. In figure 5 above, frame 5, the way is blocked for the Tesla by a black car turning left. While the Tesla has 'rights' to go next (since it has been waiting at the intersection), a blue car appears on the right and slows down, waiting for a responsive acceleration by the Tesla into the intersection. But intersections 'flow' - and if a space is not taken, then another car will drive. The blue car slows as it approaches the intersection, but the Tesla fails to move, and then the blue car itself starts to accelerate. It is only now that the Tesla starts to accelerate, but now it is going simultaneous with the other car, rather than when the car was stopping to give the Tesla a chance to go. Since the Tesla missed that 'slot' the blue car continues to accelerate and the Tesla brakes because its path through the intersection is no longer free. This braking action, in turn causes a problem for its next attempt because now it is again 'producing a yield' which the next car on the intersection uses to produce a reciprocal 'go'. The production of the joint social action of yielding/going, here has an important temporal element – the Tesla cannot react 'in time' to be able to produce the joint social action that will get it through the intersection. It fails to display a 'reflexive adjustment' to the other cars motions. The problem then that the Tesla faces - as do all self-driving cars - is not one of doing a California stop or not, rather it is producing the correct response to another driver both sequentially - after the other drivers actions - but also in time - before the their non-response is interpreted as a yield.

The Tesla can 'see' other cars slowing, and project that this is a 'yield', which allows it to 'go', but it does not seem to do this in enough time. To make this worse, it cannot understand how its own stop actions may be taken by other drivers as a yield. That is to say it fails in terms of mutual understanding of the four way stop as a system of negotiated emergent traffic order - and while it does get through the intersection, it does so with the sort of slow and incompetent driving that would likely cause many problems on the roads if repeated at scale.

6 DISCUSSION

> "we focus specifically on yielding [...] We show that road users communicate and negotiate their spatial copresence and mobile trajectories by designing their mobile actions so that they are recognizable to other road users as doing yielding." [31]

In our data analysis above we have focused on a relatively straightforward element of road interaction - yielding for other road users. The introduction of self-driving cars transforms this into a human computer interaction issue – how can self-driving cars interact successfully with other road users. Traffic has become in a sense a new 'genera' of human machine interaction - one where there are very longstanding conventions, expectations, and human interactions. While the cars we have studied do manage, mostly, to drive and pass through traffic safely, they leave a trail of confusion amongst other road users, and we must wonder how safe this really is. Confusion about intent, particularly when driving at speed, could easily cascade into collisions or other problems on the road (such as holding up traffic).

In discussion we address three issues which emerge from our data analysis. We start by addressing the broad question of what joint social action is in a situation of joint movement and attention like traffic. We return to some of the fundamental questions posted in Suchman's Plans and Situated Action over thirty years ago - can asocial technologies interact and take a place in a world of joint social action? This has relevance not just for self-driving cars, but also for other computer-controlled moving agents of different sorts.

Second, we discuss how HCI should approach the design problem of this new form of human machine interaction- and how HCI can have a role in such a fast moving area, particularly where access to the actual systems is relatively constrained. Lastly, we reflect upon the nature of our data, and the ways in which users are being integrated into testing these systems (both willingly and unwillingly) and what this says about the role of researchers.

6.1 Intersubjectivity and the Schegloff/Suchman problem

> "The procedural basis for locating and dealing with breakdowns in intersubjectivity is woven into the very warp and weft of ordinary conversation and, by implication, possibly of any organized conduct." [68]

Across our four extracts a key practical problem for those involved was recognising what the self-driving car was doing, and following those actions with appropriate next actions. This is a very practical matter - I need to know what you are doing so I can cross the road in-front of you. But it is also a core building block for joint social action, where what one agent does is coupled to the actions of another. While Intersubjectivity is often treated as some sort of rather mysterious feature of human life, Schegloff instead argued (building on Garfinkel and Schutz) that our seen-in-common social world is actually held together by many mundane conversational mechanisms [69]. Intersubjectivity depends on shared practices by which we display how we understand each others' actions, by means of our own next action. A question is followed by an answer, and by producing an answer we demonstrate that the previous action was understood to be a question. Intersubjectivity rests upon a world where we more or less for all practical purposes understand each other, even in the most fleeting of encounters [43].

And so for our drivers, driving together is not just safely moving through space, it is also producing a joint inter-subjective social world - traffic. Understanding each other, and displaying that understanding by going when the other yields, is an elementary but also fundamental form of social action. It is into this world that self-driving cars are introduced, with only the most rudimentary ability to produce this social action, or none at-all. In her book "Plans and situated action" Suchman [71] raised the important question of whether joint action can ever be produced between humans and computers, when those computers have little or no understanding of our joint social world. This point is both conceptual and practical - to what extent can mutually understandable joint action be produced between humans and moving robots – be they self-driving cars, or more complex systems. Schegloff pioneered our understanding of many of the basics of sequential interaction, including how repair is a invaluable part of conversational interaction. We can fail to understand or coordinate with each other, but we can also repair those troubles - with a gesture or a comment - "sorry self-driving car" - or a gesture to a waiting pedestrian. For Schegloff repair was a fundamental 'last defence' for maintaining intersubjectivity in our joint action. If we are not understanding each other, repair in subsequent turns in talk can be used to fix and realign speakers and their joint actions. This makes our joint interactions remarkably robust despite frequent mishearing, misunderstandings and so on.

Repair is built into much of human interaction and quite fundamental in that – in spoken interaction there are a rich set of ways in which we can repair troubles in speaking, hearing or understanding what is said [47]. While on the road our options are not quite so broad, we can certainly 'deal with' incompetent driving and the trouble this produces. As our clips above show, we can resort to hand gestures, spoken comments and so on. Even if an autonomous car cannot respond to such actions (as the case of the Waymo car that has no driver), other road users can just 'wait the car out'; drive around; or jump ahead of its turn. As our videos show, other road users – other participants in traffic – have resources they can use to deal with problematic drivers. On the road alone there are cases of distracted driving, incompetent driving, learner drivers, mechanical faults, intoxicated drivers, inexperienced drivers and so on. On the road we have a fairly reliable (although not perfect) set of methods for dealing with this incompetent driving – we stop, drive around, report to the police, and so on. In these ways, human road users can manage trouble generated by others' inappropriate driving behavior.

And in our extracts it falls to the humans to fix problematic situations by resorting to gazes, staring, apologies or hand gestures. In our first and third video clips, those involved intervene to repair the interactional trouble. In both the second and fourth clips other drivers explicitly look at the self-driving car to try to see if there is any indication of what the car is going to do next. This is an attempt to see what is going on, to resolve ambiguity through the search for a gesture or other signal. We are left then to wonder, if self-driving cars cannot resort to these measures to repair their actions, what chances are there for holding onto joint social action on the road? While self-driving cars could likely adopt machine learning models to emulate and produce a veneer of social-like action, they do not have sufficient complexity to actually engage in social behaviour (such as through repair). Without being able to repair actions and misunderstandings they will remain a-social actors on the road. Moreover, as a-social actors their lack of understanding on the road will put a burden on others to both recognize these a-social cars, and to work and plan around them. This burden will fall onto *other* road users - pedestrians, cyclists and drivers without self-driving cars.

6.2 Traffic as a genre of human/machine interaction

We acknowledge that the systems we are studying here are all under rapid development, with many developers attempting to address problems with the cars' behaviour. It may well be that the next version of the Tesla software or Waymo system, may fix and address the issues that we have outlined here. As with the debates over general machine intelligence and the rapid development of machine learning systems trained on large datasets, it is hard to argue with the ability of a future system. As we mentioned above, our paper is not an evaluation of these systems forever into the future, or even a evaluation of their current state of development. Nor are we attempting to identify bugs or simple behavioural issues that can be repaired. Rather our aim is to outline more fundamental points about this new type of human/machine interaction, in particular with respect to a self-driving car's ability to socially interact with other road users, but also to point a direction for future contributions from HCI.

We have attempted here to document the movements of cars and other road users and how such actions communicate current actions and project their next actions to other road users. Traffic is the joint production of drivers (and other road users) collaborating through an awareness of others' movements, as well as how ones' own movement appears to others. With the arrival of self-driving cars this introduces a host of design challenges and potential problems. This changing nature of public space connects with our long standing interest in how complex co-operative situations are affected by new technological interventions, often when those technical systems have been designed without an understanding of a normative organised social world that is then being disturbed. If we accept social interaction as a part of successful driving practice, we then have to consider how and whether autonomous vehicles can adequately manage and deal with social situations, and to what extent self-driving cars not being able to understand social interaction. This is not only a technical question but a broader regulatory one.

We believe that the approach to data collection and analysis taken in this paper can be used as an example for how to analyse and understand human/machine co-present interaction more broadly. It can also provide clear technical design challenges: for example: to what extent can autonomous social motion be designed - and in what shape and form might it be understood (or misunderstood) in interaction? This generates a need for documenting and understand the emergent new forms of social interaction, in our case interactions on the road, but it can also be used to study the fleeting interactions between humans and moving agents, such as human robot interaction. As such we are hopeful that our analysis can open the possibility of exploring and designing for this 'fleeting [...] interactional space' [58]. Specifically we are interested in how the sequential and temporal nature of interaction that could be part of how a system reasons about its motion. For a robot to potentially make offers (such as 'you go first'), it need reason and process to some extent the social aspects of bodily interaction.

6.3 Testing

Our last point in the discussion concerns the role that testing is playing in the design of self-driving cars. Testing is an important part of any development process, but self-driving cars complicate this hugely in that testing is also an integral part of development itself. Testing provides the data that is needed for training the system, and the newly improved systems are then released and used to collect even more data. Added to this development/testing loop, in the cases we have looked at here, public videos have made testing itself a public matter; some self-driving car videos have millions of views. Public beta tests are hardly new - but is seems unusual for a CEO to direct his developers to address the problem of one specific YouTube-posting user. In this way YouTube video content ends up guiding the testing, guiding the data that is collected and ultimately the system that is produced.

One problem with this sort of 'testing in public' is that the types and form of testing are dictated at least to some extent by the desires of the audience and creators, rather than strictly the development of the system. Indeed, the testers who post their videos on YouTube spend much of their videos apologising for the system, and perhaps even explaining away the problems that arise. And while these videos do involve impressive setups and interesting video content, they are clearly something of a supplement rather than a replacement for more directed testing. It is not clear to us, for example, that being able to do a multi-lane left turn ("Chuck's corner") is the most urgent open issue for the Tesla FSD system. We might also question who participates in the testing - should technology enthusiasts literally drive technology development, and what does this mean for other road users? Cyclists and pedestrians may be just as affected by self-driving cars, but will have no direct say in what is prioritised, tested and implemented. There is also the related burden that these systems put on other road users during their testing. While the drivers and passengers have consented to being part of a tests, other road users have not. There is clearly some responsibility on system developers to not overly burden or even potentially harm other road users. While meeting these systems being tested might be just a small nuisance, perhaps the more important issue is whether such testing is actually safe, and the possibility of serious accidents is certainly something that ought to be balanced against the benefits of generating training material for future software updates.

7 CONCLUSION

The title of this paper references one of the classic fundamental theorems of computability - whether a computer can work out if its own program is faulty, and halting if it is. As our data shows, stopping for a self-driving car is not just the outcome of some computation, but also potentially a social action. In our data we explored if software-driven cars can competently stop in traffic, and how doing so relies upon responding to the actions of other drivers. This joint action is quite different from the imaginary space of computability, and we explored to what extent social joint action can arise from the interactions of drivers with self-driving cars. To build self-driving cars that can integrate in traffic, we are convinced more than ever that understanding traffic has a fundamental role to play in debates and design of self-driving vehicles. As we argue here, the challenge then is not one of computability but of understanding social interaction.

ACKNOWLEDGMENTS

Eric Stayton of Alliance Innovation Lab Silicon Valley provided invaluable insights during the videos analysis and with the connections to self-driving car design. We also thank Eric Laurier for helpful comments on the drafts and insights into ways to take the paper forward. We would also like to acknowledge the Swedish science foundation (grant RIT15-0046 "Implicit Interaction") and WASP-HS ("AI in Motion: Studying the Social World of Autonomous Vehicles", MMW 2020.0086) for financial support for this project.

REFERENCES

[1] Claudia Ackermann, Matthias Beggiato, Luka-Franziska Bluhm, and Josef Krems. 2018. Vehicle movement and its potential as implicit communication signal for pedestrians and automated vehicles. In *Proceedings of the 6th Humanist Conference*. 1–7.

[2] Mark Bergen. 2015. Meet the Companies Building Self-Driving Cars for Google and Tesla (And Maybe Apple). http://www.recode.net/2015/10/27/11620026/meet-the-companies-building-self-driving-cars-for-google-and-tesla

[3] Neal E. Boudette. 2022. Tesla Recalls Cars With Full Self-Driving to Prevent Rolling Stops. *The New York Times* (Feb. 2022). https://www.nytimes.com/2022/02/01/business/tesla-recall.html

[4] Mathias Broth, Jakob Cromdal, and Lena Levin. 2018. Showing where you're going. Instructing the accountable use of the indicator in live traffic. *International Journal of Applied Linguistics* 28, 2 (2018), 248–264.

[5] Barry Brown and Eric Laurier. 2012. The normal natural troubles of driving with GPS. ACM, 1621–1630.

[6] Barry Brown and Eric Laurier. 2017. The trouble with autopilots: Assisted and autonomous driving on the social road. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. ACM Press, 416–429.

[7] Barry Brown, Eric Laurier, and Erik Vinkhuyzen. 2023. Designing motion: Lessons for self-driving and robotic motion from human traffic interaction. *Proceedings of the ACM on Human-Computer Interaction* 7, GROUP (2023), 1–21. Publisher: ACM New York, NY, USA.

[8] Barry Brown, Moira McGregor, and Eric Laurier. 2013. iPhone in vivo: video analysis of mobile device use. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1031–1040.

[9] Cesar Cadena, Luca Carlone, Henry Carrillo, Yasir Latif, Davide Scaramuzza, Jose Neira, Ian D. Reid, and John J. Leonard. 2016. Past, Present, and Future of Simultaneous Localization And Mapping: Towards the Robust-Perception Age. *IEEE Transactions on robotics* 32, 6 (June 2016), 1309–1332. https://arxiv.org/abs/1606.05830v2

[10] Chia-Ming Chang, Koki Toda, Xinyue Gui, Stela H. Seo, and Takeo Igarashi. 2022. Can Eyes on a Car Reduce Traffic Accidents?. In *Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '22)*. Association for Computing Machinery, New York, NY, USA, 349–359. https://doi.org/10.1145/3543174.3546841

[11] Mark Colley and Rukzio Rukzio. 2020. A Design Space for External Communication of Autonomous Vehicles. In *Proceedings of the 12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI'20)*. ACM Press, 212–222. https://doi.org/10.1145/3409120.3410646

[12] Johnna Crider. 2022. Tesla began rolling out FSD Beta 10.69; solved Chuck's Unprotected Left Turn. https://www.teslarati.com/tesla-fsd-beta-10-69-chucks-left-turn/

[13] Elwys De Stefani. 2018. Formulating direction: Navigational instructions in driving lessons. *International Journal of Applied Linguistics* 28, 2 (2018), 283–303.

[14] Elwys De Stefani and Anne-Danièle Gazin. 2018. Learning to communicate: Managing multiple strands of participation in driving lessons. *Language & Communication* (2018).

[15] Arnulf Deppermann. 2018. Instruction practices in German driving lessons: Differential uses of declaratives and imperatives. *International Journal of Applied Linguistics* 28, 2 (2018), 265–282.

[16] Arnulf Deppermann, Lorenza Mondada, and Simona Pekarek Doehler. 2021. Early Responses: An Introduction. *Discourse Processes* 58, 4 (April 2021), 293–307. https://doi.org/10.1080/0163853X.2021.1877516 Publisher: Routledge _eprint: https://doi.org/10.1080/0163853X.2021.1877516.

[17] Ian Dersley and Anthony Wootton. 2000. Complaint sequences within antagonistic argument. *Research on Language and Social Interaction* 33, 4 (2000), 375–406.

[18] Debargha Dey, Azra Habibovic, Melanie Berger, Devanshi Bansal, Raymond H. Cuijpers, and Marieke Martens. 2022. Investigating the Need for Explicit Communication of Non-Yielding Intent through a Slow-Pulsing Light Band (SPLB) eHMI in AV-Pedestrian Interaction. In *Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '22)*. Association for Computing Machinery, New York, NY, USA, 307–318. https://doi.org/10.1145/3543174.3546086

[19] Debargha Dey and Jacques Terken. 2017. Pedestrian interaction with vehicles: roles of explicit and implicit communication. In *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. ACM, 109–113.

[20] Kevin Drum. 2015. Was the California Stop really invented in California? *Mother Jones* (2015). https://www.motherjones.com/kevin-drum/2015/10/was-california-stop-really-invented-california/

[21] The Economist. 2019. Driverless cars are stuck in a jam. *The Economist* (2019). https://www.economist.com/leaders/2019/10/12/driverless-cars-are-stuck-in-a-jam

[22] Amir Efrati, Profile archive →, , and Follow Amir on Twitter. 2018. Waymo's Big Ambitions Slowed by Tech Trouble. https://www.theinformation.com/articles/waymos-big-ambitions-slowed-by-tech-trouble

[23] Jennifer Elias. 2022. Taking a driverless Waymo in Phoenix over the holidays was fun but unsettling. https://www.cnbc.com/2022/01/08/heres-what-it-was-like-to-ride-in-a-waymo-with-no-driver-in-phoenix.html

[24] Hadas Erel, Tzachi Shem Tov, Yoav Kessler, and Oren Zuckerman. 2019. Robots are Always Social: Robotic Movements are Automatically Interpreted as Social Cues. In *Extended abstracts of the 2019 CHI conference on human factors in computing systems*. 1–6.

[25] Stefanie M. Faas, Andrea C. Kao, and Martin Baumann. 2020. A longitudinal video study on communicating status and intent for self-driving vehicle–pedestrian interaction. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–14.

[26] Grace de la Flor, Paul Luff, Marina Jirotka, John Pybus, Ruth Kirkham, and Annamaria Carusi. 2010. The Case of the Disappearing Ox: Seeing Through Digital Images to an Analysis of Ancient Texts. ACM, 1753397, 473–482. https://doi.org/10.1145/1753326.1753397

[27] Andrew Gambino and S. Shyam Sundar. 2019. Acceptance of Self-Driving Cars: Does Their Posthuman Ability Make Them More Eerie or More Desirable?. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3290607.3312870

[28] Harold Garfinkel. 1967. *Studies in Ethnomethodology.* Prentice Hall.

[29] Azra Habibovic, Victor Malmsten Lundgren, Jonas Andersson, Maria Klingegård, Tobias Lagström, Anna Sirkka, Johan Fagerlönn, Claes Edgren, Rikard Fredriksson, Stas Krupenia, Dennis Saluäär, and Pontus Larsson. 2018. Communicating Intent of Automated Vehicles to Pedestrians. *Frontiers in Psychology* 9 (Aug. 2018). https://doi.org/10.3389/fpsyg.2018.01336

[30] Pentti Haddington. 2012. Movement in action: Initiating social navigation in cars. *Semiotica* 2012, 191 (Jan. 2012), 137–167. https://doi.org/10.1515/sem-2012-0059

[31] Pentti Haddington and Mirka Rauniomaa. 2014. Interaction between road users: offering space in traffic. *Space and culture* 17, 2 (2014), 176–190.

[32] Dave Harley and Geraldine Fitzpatrick. 2009. YouTube and intergenerational communication: the case of Geriatric1927. *Universal Access in the Information Society* 8, 1 (April 2009), 5–20. https://doi.org/10.1007/s10209-008-0127-y

[33] Christian Heath, Jon Hindmarsh, and Paul Luff. 2010. *Video in qualitative research: analysing social interaction in everyday life.* Sage, London.

[34] Christian Heath and Paul Luff. 2000. *Technology in action.* Cambridge university press.

[35] Nicholas J. Hetherington, Ryan Lee, Marlene Haase, Elizabeth A. Croft, and HF Machiel Van der Loos. 2021. Mobile Robot Yielding Cues for Human-Robot Spatial Interaction. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 3028–3033.

[36] Juan Pablo Hourcade, Sarah L. Mascher, David Wu, and Luiza Pantoja. 2015. Look, My Baby Is Using an iPad! An Analysis of YouTube Videos of Infants and Toddlers Using Tablets. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 1915–1924. https://doi.org/10.1145/2702123.2702266

[37] N. Jones and G. Raymond. 2012. "The Camera Rolls": Using Third-Party Video in Field Research. *The ANNALS of the American Academy of Political and Social*

Science 642, 1 (July 2012), 109–123. https://doi.org/10.1177/0002716212438205
[38] Wendy Ju. 2015. *The design of implicit interactions*. Vol. 8. Morgan & Claypool Publishers. http://www.morganclaypool.com/doi/abs/10.2200/S00619ED1V01Y201412HCI028
[39] Hermann Kaindl, Jürgen Falb, and Cristian Bogdan. 2008. Multimodal communication involving movements of a robot. In *CHI'08 Extended Abstracts on Human Factors in Computing Systems*. ACM, 3213–3218.
[40] Young Woo Kim, Jae Hyun Han, Yong Gu Ji, and Seul Chan Lee. 2020. Exploring the Effectiveness of External Human-Machine Interfaces on Pedestrians and Drivers. In *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. 65–68.
[41] Xiaoqiang Kong, Subasish Das, Yunlong Zhang, and Xiao Xiao. 2021. Lessons learned from pedestrian-driver communication and yielding patterns. *Transportation research part F: traffic psychology and behaviour* 79 (2021), 35–48. Publisher: Elsevier.
[42] Eric Laurier. 2014. The Graphic Transcript: Poaching Comic Book Grammar for Inscribing the Visual, Spatial and Temporal Aspects of Action. *Geography Compass* 8, 4 (2014), 235–248. https://doi.org/10.1111/gec3.12123 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/gec3.12123.
[43] Eric Laurier, Barry Brown, and Lorimer Hayden. 2012. What it means to change lanes: Actions, emotions and wayfinding in the family car. *Semiotica* 2012, 191 (Jan. 2012). https://doi.org/10.1515/sem-2012-0058
[44] Eric Laurier and Hayden Lorimer. 2009. Other ways: landscapes of commuting. *Landscape Research* 37, 2 (2009), 207–224.
[45] Eric Laurier, Hayden Lorimer, Barry Brown, Owain Jones, Oskar Juhlin, Allyson Noble, Mark Perry, Daniele Pica, Philippe Sormani, Ignaz Strebel, and others. 2008. Driving and 'passengering': Notes on the ordinary organization of car travel. *Mobilities* 3, 1 (2008), 1–23. http://www.tandfonline.com/doi/abs/10.1080/17450100701797273
[46] Yee Mun Lee, Ruth Madigan, and Jorge Garcia. 2019. Understanding the Messages Conveyed by Automated Vehicles. In *Proceedings of AutoUI 2019*. 134–143. https://doi.org/10.1145/3342197.3344546
[47] Gene Lerner and Celia Kitzinger. 2007. Extraction and aggregation in the repair of individual and collective self reference. *Discourse Studies* 9, 4 (2007), 526–557.
[48] Stephanie Lowry, Niko Sunderhauf, Paul Newman, John J. Leonard, David Cox, Peter Corke, and Michael J. Milford. 2016. Visual Place Recognition: A Survey. *IEEE Transactions on Robotics* 32, 1 (Feb. 2016), 1–19. https://doi.org/10.1109/TRO.2015.2496823
[49] Karthik Mahadevan, Sowmya Somanath, and Ehud Sharlin. 2018. Communicating Awareness and Intent in Autonomous Vehicle-Pedestrian Interaction. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, 429:1–429:12. https://doi.org/10.1145/3173574.3174003 event-place: Montreal QC, Canada.
[50] G. Markkula, R. Madigan, D. Nathanael, E. Portouli, Y. M. Lee, A. Dietrich, J. Billington, A. Schieben, and N. Merat. 2020. Defining interactions: a conceptual framework for understanding interactive behaviour in human and automated road traffic. *Theoretical Issues in Ergonomics Science* 21, 6 (Nov. 2020), 728–752. https://doi.org/10.1080/1463922X.2020.1736686 Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/1463922X.2020.1736686.
[51] Markus Maurer, J. Christian Gerdes, Barbara Lenz, and Hermann Winner. 2016. *Autonomous driving: technical, legal and social aspects*. Springer. http://dl.acm.org/citation.cfm?id=2965082
[52] Patrick McGee. 2021. Rolling out driverless cars is 'extraordinary grind', says Waymo boss. *Financial Times* (Jan. 2021).
[53] Paul McIlvenny, Mathias Broth, and Pentti Haddington. 2014. Moving together: Mobile formations in interaction. *Space and Culture* 17, 2 (2014), 104–106.
[54] Natasha Merat, Yee Mun Lee, Gustav Markkula, Jim Uttley, Fanta Camara, Charles Fox, André Dietrich, Florian Weber, and Anna Schieben. 2019. How Do We Study Pedestrian Interaction with Automated Vehicles? Preliminary Findings from the European interACT Project. In *Road Vehicle Automation 6 (Lecture Notes in Mobility)*, Gereon Meyer and Sven Beiker (Eds.). Springer International Publishing, 21–33.
[55] Sara Merlino and Lorenza Mondada. 2018. Crossing the street: How pedestrians interact with cars. *Language & Communication* 65 (2018), 131–147.
[56] Cade Metz. 2021. The Costly Pursuit of Self-Driving Cars Continues On. And On. And On. *The New York Times* (May 2021). https://www.nytimes.com/2021/05/24/technology/self-driving-cars-wait.html
[57] Cade Metz. 2021. Tesla Sells 'Full Self-Driving,' but What Is It Really? *The New York Times* (Aug. 2021). https://www.nytimes.com/2021/08/20/technology/tesla-full-self-driving-fsd.html
[58] Lorenza Mondada. 2009. Emergent focused interactions in public places: A systematic analysis of the multimodal achievement of a common interactional space. *Journal of pragmatics* 41, 10 (2009), 1977–1997. Publisher: Elsevier.
[59] Lorenza Mondada. 2018. Driving instruction at high speed on a race circuit: Issues in action formation and sequence organization. *International Journal of Applied Linguistics* 28, 2 (2018), 304–325.
[60] Lorenza Mondada. 2018. Multiple Temporalities of Language and Body in Interaction: Challenges for Transcribing Multimodality. *Research on Language and Social Interaction* 51, 1 (Jan. 2018), 85–106. https://doi.org/10.1080/08351813.2018.1413878 Publisher: Routledge _eprint: https://doi.org/10.1080/08351813.2018.1413878.
[61] Lorenza Mondada. 2021. How Early can Embodied Responses be? Issues in Time and Sequentiality. *Discourse Processes* 58, 4 (April 2021), 397–418. https://doi.org/10.1080/0163853X.2020.1871561 Publisher: Routledge _eprint: https://doi.org/10.1080/0163853X.2020.1871561.
[62] Lars Müller, Malte Risto, and Colleen Emmenegger. 2016. The Social Behavior of Autonomous Vehicles. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct (UbiComp '16)*. ACM, New York, NY, USA, 686–689. https://doi.org/10.1145/2968219.2968561 event-place: Heidelberg, Germany.
[63] Amir Rasouli and John K. Tsotsos. 2019. Autonomous vehicles that interact with pedestrians: A survey of theory and practice. *IEEE Transactions on Intelligent Transportation Systems* 21, 3 (2019), 900–918.
[64] Jakob Reinhardt, Lorenz Prasch, and Klaus Bengler. 2021. Back-off: Evaluation of Robot Motion Strategies to Facilitate Human-Robot Spatial Interaction. *ACM Transactions on Human-Robot Interaction (THRI)* 10, 3 (2021), 1–25. Publisher: ACM New York, NY, USA.
[65] Lindsay Reynolds, Samantha Gillette, Jason Marder, Zachary Miles, Pavel Vodenski, Ariella Weintraub, Jeremy Birnholtz, and Jeff Hancock. 2011. Contact stratification and deception: blackberry messenger versus SMS use among students. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work*. ACM, 221–224. http://dl.acm.org/citation.cfm?id=1958857
[66] Malte Risto, Colleen Emmenegger, Erik Vinkhuyzen, Melissa Cefkin, and Jim Hollan. 2017. Human-Vehicle Interfaces: The Power of Vehicle Movement Gestures in Human Road User Coordination. In *Driving Assesment Conference*, Vol. 9. University of Iowa. tex.ids: risto_human-vehicle_2017.
[67] A. Lincoln Ryave and James N. Schenkein. 1974. Notes on the art of walking. In *Ethnomethodology*, R. Turner (Ed.). Penguin, Harmondsworth, 265–74.
[68] Emanuel A. Schegloff. 1992. Repair After Next Turn: The Last Structurally Provided Defense of Intersubjectivity in Conversation. *Amer. J. Sociology* 97, 5 (1992), 1295–1345. https://www.jstor.org/stable/2781417
[69] Emanuel A. Schegloff. 2007. *Sequence Organization in Interaction: Volume 1: A Primer in Conversation Analysis*. Cambridge University Press.
[70] David Silverman. 1998. *Harvey Sacks: social science and conversation analysis*. Polity Press, Cambridge.
[71] Lucy Suchman. 2007. *Human-machine reconfigurations: plans and situated actions* (2nd ed ed.). Cambridge University Press, Cambridge ; New York. OCLC: ocm64592145 tex.ids= suchman_human-machine_2007-1.
[72] Sebastian Thrun, Michael Montemerlo, and Andrei Aron. 2006. Probabilistic Terrain Analysis For High-Speed Desert Driving.. In *Robotics: Science and Systems*. 16–19. http://robots.stanford.edu/papers/thrun.mapping-Stanley.pdf
[73] Chris Urmson, Joshua Anhalt, Drew Bagnell, Christopher Baker, Robert Bittner, M. N. Clark, John Dolan, Dave Duggins, Tugrul Galatali, Chris Geyer, and others. 2008. Autonomous driving in urban environments: Boss and the urban challenge. *Journal of Field Robotics* 25, 8 (2008), 425–466. https://doi.org/10.15607/RSS.2006.II.021
[74] Remo M.A. van der Heiden, Shamsi T. Iqbal, and Christian P. Janssen. 2017. Priming Drivers Before Handover in Semi-Autonomous Cars. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 392–404. https://doi.org/10.1145/3025453.3025507 event-place: Denver, Colorado, USA.
[75] Marcel Walch, Kristin Lange, Martin Baumann, and Michael Weber. 2015. Autonomous Driving: Investigating the Feasibility of Car-driver Handover Assistance. In *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '15)*. ACM, New York, NY, USA, 11–18. https://doi.org/10.1145/2799250.2799268 event-place: Nottingham, United Kingdom.
[76] Rod Watson. 1999. Driving in forests and mountains: A pure and applied ethnography. *Ethnographic studies* 3 (1999), 50–60. https://doi.org/10.5449/idslu-001103654