Reflections on the relationship between algorithmic transparency and the social acceptability of autonomous vehicles.

Tommaso Colombino
Naver Labs Europe
Meylan, France

tommaso.colombino@naverlabs.com

ABSTRACT

This position paper discusses the relationship between the social acceptability of autonomous and semi-autonomous vehicles and the transparency of the underlying artificial intelligence algorithms that drive autopilot technology. I draw on existing, ethnographically informed approaches to the design of interface / interaction mechanisms and autopilot features that advocate for the need to make autonomous vehicles appear more competent with respect to the social and organizational properties of traffic, and argue that this approach may come at the cost of transparency, and the opportunity for users to usefully and successfully appropriate the technology.

Workshop proceedings *Everyday Automation Experience'19*
In conjunction with CHI'19, May 5th, 2019, Glasgow, UK
Website: http://everyday-automation.tech-experience.at

KEYWORDS

Autonomous vehicles; ethnography; interface design; artificial intelligence; interaction design

INTRODUCTION

The pervasiveness and growing impact (and decision making power) of AI and algorithmic technologies has led researchers in the fields of social studies of science and technology, HCI and CSCW to adopt the notion of "nonhuman agency" [2], and to promote "empirical investigations of the concrete practices through which categories of human and nonhuman are mobilized and become salient within particular fields of action" [11]. While AI research has evolved greatly in the last 30 years and moved away somewhat from the rule-based systems Suchman originally studied, it can also be argued that the goal of modern algorithm-driven systems (be it conversational agents, recommender systems, and more generally user interfaces to systems with complex underlying algorithms) remains the same: creating a blackboxed [6] service or infrastructure with an underlying technology that remains largely hidden to the user. Or to put it less critically, as is done for example when describing the underlying design principles of ambient intelligence, user interaction design for AI-driven systems aims to provide an experience where physical and digital artefacts integrate seamlessly with and effectively disappear or sublimate into the service they provide [9]. Consequently, it could be argued that the challenges faced in designing such systems for users (or for usability) remain as salient as they were three decades ago.

Autonomous and semi-autonomous vehicles, in this context, represent a particularly rich example. Cars are ubiquitous and increasingly integrated in and with technology, and automation would almost seem like a logical next step from driver assistance technologies which are now commonplace (like ABS, traction control, automatic emergency braking, etc.). And traffic and driving are experiences common to most people, but rife with the kind of social nuance and complexity that are likely to pose serious challenges to automation.

In the Association for Computing Machinery, Computer Supported Co-operative Work and Human Computer Interaction research communities, attention has been paid to topics such as: how autonomous vehicles may communicate amongst each other [3]; how fleets of autonomous vehicles might be managed [7]; the haptics and controls of autonomous and semi-autonomous vehicles [5]; and more recently, researchers have started to pay attention to communication of intent between pedestrians and autonomous vehicles [8][4].

With driverless cars being a relatively new phenomenon, there are perhaps fewer observational or ethnographic studies of driver / vehicle or driver / pedestrian interactions (see for example [9] on how drivers behave in semi-autonomous vehicles and on how pedestrians and driverless cars might communicate or signal their intention to cross the road). Unsurprisingly, car manufacturers with an interest in developing autonomous vehicles have taken to hiring social scientists and ethnographers to study things like how traffic is regulated through codified rules and laws (which would be relatively easy to teach to an "intelligent" system), but is also obviously a social and emergent phenomenon where rights and opportunities are constantly (re)negotiated through the behavior of agents on the road.

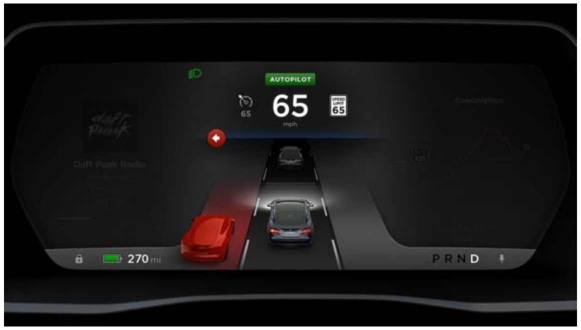
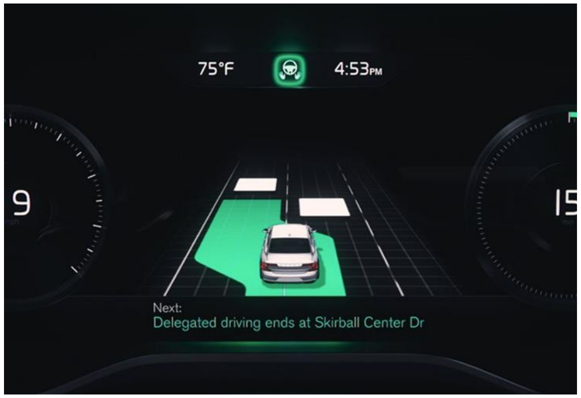


Figure 1: Tesla Interface



Next:
Delegated driving ends at Skirball Center Dr

Figure 2: Volvo Interface

BACKGROUND & CONTEXT

The clearest agenda for a systematic ethnographic exploration of traffic as an emergent phenomenon to inform the development of autonomous vehicle technology is laid out by [1]. In their study of Youtube videoclips of largely uncut footage of autopilot features being used in traffic, they bring together concerns about how the behavior of an AI agent might be interpreted by users inside and outside the vehicle within the broader social context of traffic, and what kind of information about its underlying mechanisms (that is to say, what degree of transparency) would facilitate the use of and interaction(s) with an autonomous or semi-autonomous vehicle.

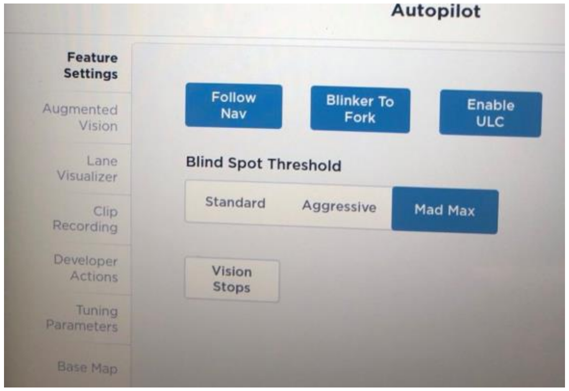
In their conclusion Brown and Laurier suggest that the way forward from a design point of view would be to move from a traditional concept of transparency (a representation of an internal state of the AI), towards making the actions undertaken by the AI "legible" as behavior in context in social terms.

A cursory examination of the interfaces of current commercially available semi-autonomous vehicles illustrates what Brown and Laurier refer to when they talk about visualizations that are reduced to simply providing information about the "state" of the AI.

The information displayed in the two very similar interfaces in Figures 1 & 2 are schematic representations of what is perceived by the AI's visual processing engine: road markings and boundaries and other vehicles. The act of changing lanes is reduced to a problem of speed and trajectories of moving objects to calculate available space and those calculations are necessary to avoid collisions. But clearly there is more to reasoning about the behavior of other drivers in traffic than calculating trajectories and available space. As Brown and Laurier point out, for example, leaving a space between yourself and the vehicle in front of you is a safety protocol to an autopilot, but can also be interpreted as a deliberate invitation to occupy that space or merge into your lane in front of you by other drivers. Their conclusion is that it might ultimately be better to design auto-pilots to exhibit behaviors (or as they say, develop movement protocols) that make the vehicle's intent more obvious.

A more practical discussion of how this might be implemented is provided by Erik Vinkhuyzen and Melissa Cefkin [12] in their description of their efforts to use detailed ethnographic descriptions of specific social practices in traffic to make algorithmic decision-making be, or appear to be, more socially competent. They outline the challenge involved in reducing social, indexical and emergent decision context into algorithmic decision making capable of generalizing across different instances of a social behavior. A potential solution that emerges from their work is to actually use the ethnographic studies to identify contingent behaviours in traffic that may fall outside of official traffic rules but by their nature can nevertheless be reduced to a set of rules and a decision logic exploitable by an AI. The example they give is that of "piggybacking" which is broadly described as a common, socially competent, and more importantly, socially acceptable way of breaking what would otherwise be a pre-established order of priorities for crossing an intersection.

Teaching an autopilot to "piggyback" should then, in principle, allow an autonomous vehicle to better integrate itself with human and fully socially competent drivers and constitute an instance of an AI competently engaging in an emergent social practice aimed at facilitating the general flow of traffic. As mentioned above, the advantage of piggybacking is that while it may be viewed as socially competent behavior it can be reduced to procedure and a calculation of available space. In other words, its socially competent side ("breaking the law" to expedite traffic flow) is an illusion. A next possible step Vinkhuyzen and Cefkin suggest to extend the illusion would be to develop an external interface that explicitly communicates some form of intent (for example to yield to a pedestrian) to other (human) agents in traffic, provided the AI is able to "read" the relevant social content. This could be an ingenious practical workaround to what otherwise would be an intractable "semantic gap" type problem. But it does also introduce problems of its own and in the remainder of this position paper I would like to focus on one which I think is particularly salient, and which goes back to Suchman's "asymmetry" between humans and machines.



Blind Spot Threshold: Standard Aggressive [✓ Mad Max]

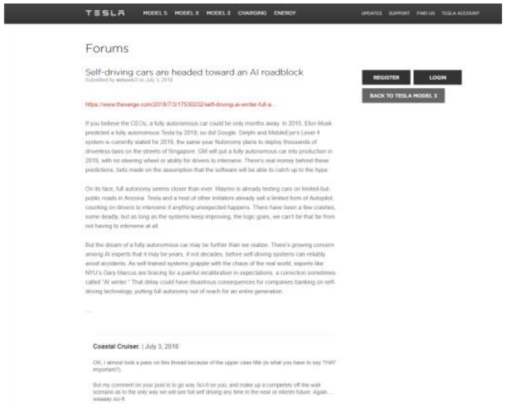
Figure 3: Elon Musk tweet on prototype autopilot features

https://twitter.com/elonmusk/status/1011021456526987266

DISCUSSION & CONCLUSION

One of the more interesting problems identified by Suchman was that the breakdown in the interactions between the photocopiers in question and their users was often instantiated by what could, as in the Piggybacking example above, be described as the illusion of social competence given in Suchman's case by a complex, expert-system based interface. We could otherwise say that the illusion of social competence brings with it the normative expectations (and therefore potentially also the sanctions) of being a socially competent agent. A transparently dumb technology comes with its own challenges but its failings are (arguably) less likely to be perceived as social infractions.

Traffic as a social setting is not just socially and normatively complex, but also particularly emotionally charged, and the agents involved in traffic (autonomous or otherwise) are likely to be perceived not just in terms of their intent but also the quality of their character. Figure 3 was tweeted by Elon Musk to provide a preview of some possible upcoming autopilot settings for Tesla's prototype autonomous readers be unfamiliar with the movies) is presented by Musk as a corrective to the fact that autopilots are easily "taken advantage of" in traffic. Obviously the concern is not for the AI's feelings but for the disruption a vehicle that slavishly and incompetently follows all the rules can cause. And the setting



Forums

Self-driving cars are headed toward an AI roadblock
Submitted by [illegible] on July 3, 2018

https://www.theverge.com/2018/7/3/17530232/self-driving-ai-winter-full-a...

If you believe the CEOs, a fully autonomous car could be only months away. In 2015, Elon Musk predicted a fully autonomous Tesla by 2018; so did Google. Delphi and MobileEye's Level 4 system is currently slated for 2019, the same year Nutonomy plans to deploy thousands of driverless taxis on the streets of Singapore. GM will put a fully autonomous car into production in 2019, with no steering wheel or ability for drivers to intervene. There's real money behind these predictions, bets made on the assumption that the software will be able to catch up to the hype.

On its face, full autonomy seems closer than ever. Waymo is already testing cars on limited-but-public roads in Arizona. Tesla and a host of other imitators already sell a limited form of Autopilot, counting on drivers to intervene if anything unexpected happens. There have been a few crashes, some deadly, but as long as the systems keep improving, the logic goes, we can't be that far from not having to intervene at all.

But the dream of a fully autonomous car may be further than we realize. There's growing concern among AI experts that it may be years, if not decades, before self-driving systems can reliably avoid accidents. As self-trained systems grapple with the chaos of the real world, experts like NYU's Gary Marcus are bracing for a painful recalibration in expectations, a correction sometimes called "AI winter." That delay could have disastrous consequences for companies banking on self-driving technology, putting full autonomy out of reach for an entire generation.

—

Coastal Cruiser. | July 3, 2018

OK, I almost took a pass on this thread because of the upper case title (is what you have to say THAT important?).

But my comment on your post is to go way Sci-fi on you, and make up a completely off-the-wall scenario as to the only way we will see full self driving any time in the near or interim future. Again... waaaay sci-fi.

Figure 4: Tesla discussion forum on AI
https://forums.tesla.com/forum/forums/self-driving-cars-are-headed-toward-ai-roadblock

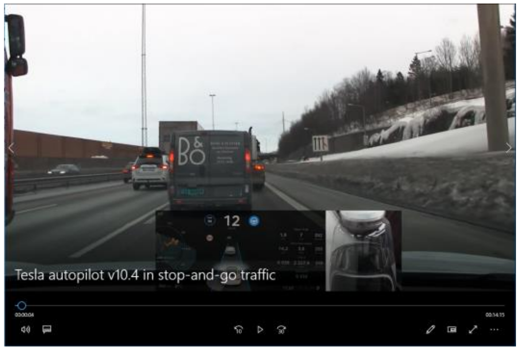


Figure 5: downloaded Youtube video demonstrating the use of the autopilot in stop-and-go traffic:
https://www.youtube.com/watch?v=aZaxfB8Lr8g

Other instances:
https://www.youtube.com/watch?v=oxrtisXUUBk
https://www.youtube.com/watch?v=_YsIQSgISj8

may be more of an inside developers' joke that a serious design proposal, but it nevertheless underlines the assumption that autonomous vehicles somehow need to engage with the normative and emotive content of traffic. It also, perhaps innocently, underlines the fact that emotions in traffic can easily turn toxic, and it raises the question of whether trying to push users away from viewing AI agents in traffic as being more socially competent (and therefore more socially accountable) that they actually are might not be an equally sensible interaction design paradigm.

As autonomous and semi-autonomous vehicles become more common, and opportunities for ethnographic studies become more widely available, it might be sensible to look at how users are appropriating the technology based on their understanding of its capabilities. Taking advantage of a naïve AI agent in traffic is just one side of the appropriation coin. The other, given a sufficient understanding of the underlying mechanisms of the technology, could be the useful and creative appropriation and integration of autopilot features into real driving practices, scenarios and routines. Tesla users appear to be particularly engaged with the emerging potential of the technology. For some the interest lies mainly in the fact that Teslas are electric vehicles, but there is also community of users that is quite visible on social media (uploading Youtube videos and participating in discussion forums like Teslarati) who are specifically interested in the autopilot features. One example that I think is pertinent comes from a user who "appropriated" the autopilot's feature of maintaining a minimum distance from cars in front and behind to get the vehicle to effectively drive itself in heavy, stop-and-go traffic. So a safety mechanism with no aspirations of being socially competent can be in fact exploited to competently negotiate a fairly common (and tedious) traffic scenario. And it is understanding the underlying mechanism such as it is (as well as transparent communication on the part of Tesla that the feature was included in a software update to the AI) that allows the user to imagine ways of making the car more useful and more competent in real traffic scenarios.

Without making overly general claims based on one example, it might nevertheless be interesting to consider that users themselves might be in a good position to do some of the work of making the technology more socially competent and accessible. In order to do that however, they have to be able to appropriate the technology and that process might be better served by interaction mechanisms and interfaces that lean more towards transparency into the underlying algorithms, and less towards sugar-coating the asymmetry between the users' and the AI's understanding of the social properties of traffic with an illusion of social competence. It may also turn out to be a more practical solution that engaging in a potentially endless exercise of trying to identify and inventory road user practices across all the possible and different operational domains, countries and cultures where automated vehicles can and eventually will be deployed.

REFERENCES

[1] Brown, Barry and Laurier, Eric (2017). The trouble with autopilots: Assisted and autonomous driving on the social road. CHI 2017, Denver, USA.

[2] Casper, Monica (1998). Reframing and grounding nonhuman agency: What makes a fetus an agent? American Behavioral Scientist 37:839-856.

[3] Driggs-Campbell, K., Shia, V. and Bajcsy, R. (2014). Decisions for Autonomous Vehicles: Integrating Sensors, Communication, and Control. HiCoNS'14, Berlin, Germany.

[4] Fingas, J. (2017, Sept. 14). Ford wants self-driving cars to communicate with flashing lights. Retrieved from https://www.engadget.com/2017/09/14/ford-self-driving-car-light-signals/

[5] Large, D., Banks, V., Burnett, G. and Margaritis, N. (2017). Putting the Joy in Driving: Investigating the Use of a Joystick as an Alternative to Traditional Controls within Future Autonomous Vehicles. Automotive UI'17, Oldenburg, Germany.

[6] Latour, Bruno (1999). Pandora's hope: essays on the reality of science studies. Cambridge, Massachusetts: Harvard University Press.

[7] Look, G. and Shrobe, H. (2004). A Plan-Based Mission Control Centre For Autonomous Vehicles. IUI'04, Madeira, Portugal.

[8] Mahadevan, Karthik, Somanath, S. and Sharlin, E. (2018). Communicating Awareness and Intent in Autonomous Vehicle-Pedestrian Interaction. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, ACM.

[9] Marshall, A. (2017, Nov. 20). MIT looks at how humans sorta drive in sorta self-driving cars. Retrieved from https://www.wired.com/story/mit-humans-semiautonomous-car-study/

[10] Stephanidis, Constantine (2011). Natural Interaction in Ambient Intelligence Environments. Keynote talk at SIGDOC'11, October 3--5, 2011, Pisa, Italy.

[11] Suchman, Lucy (2007). Human-Machine Reconfigurations: Plans and Situated Actions. Cambridge: Cambridge University Press.

[12] Vinkhuyzen, Erik and Cefkin, Melissa (2016). Developing Socially Acceptable Autonomous Vehicles. 2016 Ethnographic Praxis in Industry Conference Proceedings. https://www.epicpeople.org