Action and intention recognition in human interaction with autonomous vehicles

Maria Nilsson
Viktoria Swedish ICT
Cooperative Systems Group
Lindholmspiren 3A
41756 Göteborg, Sweden
maria.nilsson@viktoria.se

Tom Ziemke
Interaction Lab
School of Informatics
University of Skövde
54128 Skövde, Sweden
tom.ziemke@his.se

Serge Thill
Interaction Lab
School of Informatics
University of Skövde
54128 Skövde, Sweden
serge.thill@his.se

Abstract
We introduce the notion of mutual action and intention recognition in human interaction with autonomous vehicles, highlighting its relevance in successful autonomous vehicle UX. We also discuss insights and methodologies from theories of embodied cognition and social interaction relevant to these concepts. The work presented here is part of a new Swedish research initiative - AIR - whose focus is on mutual action and intention recognition in human interaction with autonomous systems in general, and automated vehicles in particular.

Author Keywords
action recognition, intention recognition, embodied cognition, social interaction

ACM Classification Keywords
H.5.2 [Information interfaces and presentation (e.g. HCI)]: User Interfaces.

Introduction
AIR is a new Swedish research initiative (2015 – 2019) on action and intention recognition in human interaction with autonomous systems moving in shared physical spaces. It is a collaborative effort led by the Interaction Lab at the University of Skövde together with researchers at the Viktoria Swedish ICT research institute, Örebro University,

and Halmstad University. Research scenarios in the AIR environment address human interaction with (a) social/assistive robots in the home, (b) autonomous transport vehicles in industrial environments, and (c) autonomous vehicles in public traffic. Here, we focus on the latter research direction.

Autonomous (driverless) cars have been a focus for research in numerous projects over the years. The focus predominantly has been on technical aspects such as sensor systems and on autonomous control systems (as several early research efforts demonstrate, *e.g.* the pan European EUREKA – PROMETHEUS project (1987-1995), the DARPA Grand and Urban Challenges, Google's driverless car, and the VisLab Intercontinental Autonomous Challenge all focus on the technical aspects of autonomous driving).

Technological progress towards highly automated vehicles is therefore well under way. However, this also creates new human-machine interaction (HMI) challenges that need to be addressed prior to a full deployment of such vehicles. Consequently, there is an increasing amount of research into understanding the issues involved in human-vehicle interaction in automated driving. These have been investigated in various European projects such as SARTE, HAVIT, CityMobile, INTERACTIVE, and V-charge.

Action recognition refers to the ability to perceive *observable*, goal-directed movements of another agent.

Intention recognition relates to inferring the *non-obseravable* motivations underlying an action, and possibly predicting future actions.

One major concern with fully automated vehicles is the design of the interaction between the user and the automatic system/vehicle. For instance, the driver (for legislative and societal reasons) likely has to remain in control of the system's safety [6]. The driver must thus either perform the driving (assistance) or monitor the automated system and take over in cases of system limitations or system failures (semi-automated up to highly automated driving). How such interaction should be designed remains an open challenge. Other open challenges regard the perceptions and actions of road users as well as their adaptation to automatic vehicles and their ability to adjust and predict the manoeuvres of such vehicles [6].

In AIR, we highlight the importance of *mutual recognition of actions and intentions* in addressing these challenges. How, for instance, does a pedestrian at a zebra crossing infer whether or not it is the intention of an approaching automated vehicle to stop and let the pedestrian cross? How can the pedestrian know that the vehicle has noticed their presence? Conversely, how does the vehicle infer whether the intention of the pedestrian is to cross the road immediately or wait until the vehicle has passed?

Mutual action and intention recognition

Research challenges

The development of autonomous vehicle UX during the 4-year effort of AIR is driven by the following questions:

1. How can one predict the actions/intentions of automated vehicles and how can automated vehicles and human behaviour be described within the traffic system?
2. What information needs to be presented to communicate the action and intentions of automated vehicles and how should it be presented? For instance, to what degree do agents need to understand each other's abilities, opportunities and limitations to achieve a mutual understanding?

On the technical side of things, there already are partial solutions addressing the challenges these questions imply. It is, for instance, already possible to recognise the presence of humans and obstacles in traffic. Algorithms to classify vehicle behaviour by modelling its behavioural states and the ability to predict state transitions including trajectory prediction are also already being developed (*e.g.* [2, 8, 4]). Future cooperative systems will further extend the perceivable environment beyond the vehicle's own perceptual field. It is therefore reasonable to expect that technological limitations will not pose a major obstacle to addressing the challenges in the coming years.

Embodied accounts of action and intention recognition
From a UX perspective, however, we also need to turn to our understanding of human cognition as it applies to interaction with autonomous systems, in particular embodied mechanisms of action and intention recognition. Research in the cognitive sciences, not least social neuroscience, has in the last 10-20 years made substantial progress in elucidating the bodily and sensorimotor mechanisms in human-human social interactions and in developing computational models thereof [12]. When it comes to mechanisms underlying the human interpretation of the behaviour of artifacts, such as robots or automated vehicles, and the attribution of intentions to such systems, however, there is much less research. Existing efforts suggest two rather different cognitive mechanisms are likely to be relevant to understanding human embodied cognition and social interaction with autonomous systems:

Embodied accounts of cognition emphasise the role of bodily and sensorimotor mechanisms in human thinking [11].

- People tend to interpret moving inanimate objects in terms of more or less human-like actions and intentions [7] and some research suggests this extends to robots [5]; it is therefore plausible that this will also apply to autonomous systems, such as automated cars.
- Much recent (social) neuroscience research, not least the discovery of the so-called mirror (neuron) system, points to the importance of embodiment and morphological similarities in social interactions [12]. This suggests that humans might be able to more or less easily understand the behaviour of very human-like robots, but not necessarily the behaviour of, for example, autonomous lawnmowers or automated vehicles.

Mirror neurons are neurons that fire both when an agent carries out an action or observes the same action [12].

For research on human interaction with automated vehicles, particularly when it concerns action and intention recognition, it is therefore important to clarify the relative importance of these mechanisms and what roles they might play specifically in this context. This points towards a need for an ability to evaluate interaction between humans and automated cars – in other words, the vehicle UX – in a manner that explicitly takes into account human embodied cognition.

Evaluating autonomous vehicle UX
There are two branches of cognitive science that are of particular interest for the evaluation of autonomous vehicle UX. The first branch contains extended and distributed views of cognition [9, 1], according to which autonomous system becomes an extension of the human mind while cognition is distributed and should be understood in terms of the interaction with the material and social world. An extensive review of the applications distributed cognition finds in HCI, for instance as a method "with which to understand the underlying mechanisms of the relationships between humans and computers" can be found in [10].

The second branch of interest is social cognition, in particular social interaction, which can be defined as "two or more autonomous agents co-regulating their coupling with the effect that their autonomy is not destroyed and their relational dynamics acquire an autonomy of their own" [3] (p. 441). This coupling is provided through mutual action and intention recognition, making methodologies developed in this field (for a comprehensive review, see [3]) relevant to our purposes.

Conclusions

We have introduced the notion of mutual action and intention recognition in human interaction with autonomous systems in shared physical spaces, as well as a major new Swedish research initiative (AIR) on these topics. To address the challenges posed by an explicit recognition of action and intention recognition as a factor for vehicle UX, we have briefly introduced relevant theoretical and methodological approaches from the cognitive sciences, in particular related to embodied cognition theories and mechanisms. The importance of this perspective for the study of human interaction with technology increases as the latter increases in autonomy. Autonomous vehicles are a prime example of this trend [13]. In coming years, a significant part of the work carried out as part of AIR will therefore focus on mutual action and intention recognition in interaction with autonomous vehicles.

Acknowledgements

This work is part of AIR (action and intention recognition in human interaction with autonomous systems), a SIDUS project funded by the Knowledge Foundation, Sweden.

References

[1] Clark, A., and Chalmers, D. The extended mind. *Analysis* (1998), 7–19.

[2] Dagli, I., Brost, M., and Breuel, G. Action recognition and prediction for driver assistance systems using dynamic belief networks. In *Agent Technologies, Infrastructures, Tools, and Applications for E-Services*, J. Carbonell et al., Eds., vol. 2592 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2003, 179–194.

[3] De Jaegher, H., Paolo, E. D., and Gallagher, S. Can social interaction constitute social cognition? *Trends Cogn. Sci. 14*, 10 (2010), 441 – 447.

[4] Firl, J., and Tran, Q. Probabilistic maneuver prediction in traffic scenarios. In *Proc. ECMR* (2011).

[5] Gazzola, V., Rizzolatti, G., Wicker, B., and Keysers, C. The anthropomorphic brain: The mirror neuron system responds to human and robotic actions. *NeuroImage 35*, 4 (2007), 1674 – 1684.

[6] Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y. C., Visser, E. D., and Parasuraman, R. A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors 53*, 5 (2011), 517 – 527.

[7] Heider, F., and Simmel, M. An experimental study of apparent behavior. *Am. J. Psychol.* (1944), 243–259.

[8] Hermann, A., and Desel, J. Driving situation analysis in automotive environment. In *Proc. ICVES* (Sept 2008), 216–221.

[9] Hutchins, E. *Cognition in the Wild*. MIT press Cambridge, MA, 1995.

[10] Nilsson, M. *Capturing semi-automated decision making: the methodology of CASADEMA*. PhD thesis, Örebro University, 2010.

[11] Shapiro, L. *Embodied cognition*. Routledge, 2010.

[12] Thill, S., Caligiore, D., Borghi, A. M., Ziemke, T., and Baldassarre, G. Theories and computational models of affordance and mirror systems: An integrative review. *Neurosci. Biobehav. Rev. 37*, 3 (2013), 491 – 521.

[13] Thill, S., Hemeren, P., and Nilsson, M. The apparent intelligence of a system as a factor in situation awareness. In *Proc. CogSIMA* (March 2014), 52–58.