

Anticipatory Human-Machine Interaction

Jelmer Borst^{*1}, Andreas Bulling^{*2}, Cleotilde Gonzalez^{*3}, and Nele Russwinkel^{*4}

1 University of Groningen, NL. j.p.borst@rug.nl

2 University of Stuttgart, DE. andreas.bulling@vis.uni-stuttgart.de

3 Carnegie Mellon University, US. coty@cmu.edu

4 TU Berlin, DE. nele.russwinkel@tu-berlin.de

Abstract

Even after three decades of research on human-machine interaction (HMI), current systems still lack the ability to predict mental states of their users, i.e., they fail to understand users' intentions, goals, and needs and therefore cannot anticipate their actions. This lack of anticipation drastically restricts their capabilities to interact and collaborate effectively with humans. The goal of this Dagstuhl Seminar was to discuss the scientific foundations of a new generation of human-machine systems that anticipate, and proactively adapt to, human actions by monitoring their attention, behavior, and predicting their mental states. Anticipation might be realized by using mental models of tasks, specific situations and systems to build up expectations about intentions, goals, and mental states that gathered evidence can be tested against.

The seminar provided an inter-disciplinary forum to discuss this emerging topic by bringing together – for the first time – researchers from a range of fields that are directly relevant but hitherto haven't met on this topic so far. This includes human-computer interaction, cognitive-inspired AI, machine learning, computational cognitive science, and social and decision sciences. We discussed theoretical foundations, key research challenges and opportunities, new computational methods, and future applications of anticipatory human-machine interaction.

Seminar May 16–20, 2022 – <http://www.dagstuhl.de/22202>

2012 ACM Subject Classification Human-centered computing → Collaborative and social computing; Human-centered computing → Human computer interaction (HCI); Human-centered computing → Collaborative interaction; Computing methodologies → Artificial intelligence; Computing methodologies → Machine learning

Keywords and phrases Human-Computer Interaction, Anticipation, Collaboration, Collaborative Intelligence, Human-AI Teaming, Multi-Agent Simulation, Artificial Intelligence

Digital Object Identifier 10.4230/DagRep.12.5.131

1 Executive Summary

Jelmer Borst (University of Groningen, NL, j.p.borst@rug.nl)

Andreas Bulling (University of Stuttgart, DE, andreas.bulling@vis.uni-stuttgart.de)

Cleotilde Gonzalez (Carnegie Mellon University, United States, coty@cmu.edu)

Nele Russwinkel (TU Berlin, DE, nele.russwinkel@tu-berlin.de)

License  Creative Commons BY 4.0 International license

© Jelmer Borst, Andreas Bulling, Cleotilde Gonzalez, Nele Russwinkel

There is growing interest in interdisciplinary understanding of the anticipatory processes regarding other people's actions and also one's own actions. This surge in interest is especially important given that anticipatory interaction is needed in the growing area of intelligent

* Editor / Organizer



Except where otherwise noted, content of this report is licensed under a Creative Commons BY 4.0 International license

Anticipatory Human-Machine Interaction, *Dagstuhl Reports*, Vol. 12, Issue 5, pp. 131–169

Editors: Jelmer Borst, Andreas Bulling, Cleotilde Gonzalez, Nele Russwinkel



DAGSTUHL
REPORTS

Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

systems that work closely together with human partners. Decades of research on human-machine interaction (HMI) have resulted in significant advances in theories, tools, and methods to facilitate, support, and enhance interactions between humans and computing systems. Despite the fundamental importance of HMI for our information society and numerous advances towards making interactions with machines more human-like, current systems still fall short in one core human ability – Theory of Mind (ToM). ToM allows us to attribute mental states to others and anticipate their actions; and is thus essential for us to interact naturally, effortlessly, and seamlessly.

ToM shapes how we interact with each other and is most easily observable in physical tasks, such as moving a table together. In this scenario, we rely on ToM abilities to attribute intentions to others and in turn, continuously adapt our own behaviour to accommodate the intentions of others, resulting in seamless collaboration. ToM begins to develop in early childhood. Even small children who are not able to develop the full cognitive ability of ToM can anticipate others' intentions using simple mental models of tasks and following familiar sequences of actions (e.g. Fiebich, 2018).

Deficits in ToM are closely linked to developmental disorders, such as autism, and current machines are similarly mind-blind. That is, they fail to sense users' attention and predict their intentions, and therefore lack the ability to anticipate and pro-actively adapt to users' actions. This limits machines to operating after the fact, i.e. to merely react to user input. This fundamentally limits the naturalness, efficiency, and user experience of current interactions. Imagine how difficult it would be to move a table with a robot: as the robot cannot anticipate easy and hard actions for the user (e.g., the table is heavy and needs to be rested for a moment, a corner has to be navigated for which the user has to switch her grip, etc.), it would not be a very helpful teammate. We believe anticipatory HMI has significant potential to bring systems that possess ToM to a new, exciting level of development. However, the building blocks required for anticipatory HMI are still at a very early research stage. This is in stark contrast to the large body of work on (Computational) Theory of Mind in the cognitive and neurosciences and the variety of potential applications in which artificial anticipatory behaviour could have a revolutionary and paradigm-changing influence.

With this seminar we have made a first step in bridging this gap and have discussed theoretical foundations, key research challenges and opportunities, new computational methods, as well as applications and use cases of anticipatory HMI. Research on anticipatory HMI is inherently interdisciplinary and draws from a number of fields, most notably human-computer interaction, machine learning, computer vision, computational neuroscience, computational cognitive science, privacy and security, as well as context-aware computing. Consequently, in this seminar, we have brought together junior and established researchers within these communities for the first time to explore the scientific foundations of anticipatory HMI.

Three perspectives on human-machine anticipation

In the seminar we have discussed three perspectives on human-machine anticipation.

Anticipating the intentions, beliefs, mental states and next actions of the user. This perspective focuses on the challenges, as well as possible solutions, for machine anticipation of user intentions, goals, motivations, and behaviors. This includes the underlying question of which mental states need to be modeled and anticipated in the first place. This might be different depending on the purpose of the collaboration or on the specific circumstances.

Anticipating the outcome of the machine's own actions. The ability of a machine to anticipate its own actions and outcomes is important, especially for interaction in teams and for learning purposes. Artificial agents need to have a concept of the outcome of their own actions, as that will naturally affect the collaboration with the human or other agents. In addition, in case the expected outcome of an action is not reached, they have to learn how they can achieve the expected outcome in an alternative manner.

Anticipating the outcome of collaborative work and human-machine teaming. The third perspective builds on the previous two. In human-machine teaming, a machine needs to anticipate the human partners' actions, identify problems, and develop ideas on how the partners in a team can be supported. We will discuss how artificial agents can engage in effective teamwork with humans and which specific abilities are required. We will also discuss whether we can construct artificial agents that behave indistinguishable from human agents – and whether this is even desirable.

Underlying concepts and methods

In addition to the three different perspectives on anticipatory behaviour, we have discussed the underlying concepts and computational methods.

On the conceptual side, we have discussed definitions and developed a common language. This included, for example, concepts such as cognitive state vs. mental states (which are not always concisely defined across disciplines). Furthermore, to support our cross-disciplined, multi-context approach, we have collected different forms and definitions of “anticipation”. A related but separate concept that we have discussed is what different disciplines understand by ToM and what related terms that should be differentiated?. We found that there are considerable differences of opinion on this that we formulated and exchanged.

On the methodological side, we have discussed the ways in which different domains and disciplines have their own methods, approaches, and challenges. Subsequently, groups have exchanged and discussed these collections to gain a deeper understanding of available methodological approaches. We have also talked about current approaches in modelling and cognitive science, and exchange ideas with those of us from other disciplines. Which approaches support the development of better applications? More general is the question how user behaviour can be measured. What data can or should be collected (CV, natural language, neurobiological measures, behavior,..), how can we process it most effectively, and how can we incorporate it into cognitive or computational models of the other users? Finally, we have discussed what the best way is to integrate and fuse different sources of information about the user, and how this can be integrated with the current context.

2 Table of Contents

Executive Summary

<i>Jelmer Borst, Andreas Bulling, Cleotilde Gonzalez, Nele Russwinkel</i>	131
---	-----

Overview of Talks

Anticipation for Assistive and Collaborative Human-Machine Interaction <i>Henny Admoni</i>	136
From Anticipatory to Socially-Aware User Interfaces <i>Elisabeth André</i>	137
Sensing and Modelling Human Attention towards Anticipatory Human-Computer Interaction <i>Mihai Bâce</i>	137
Anticipatory Human-Machine Interaction with Cognitive Models <i>Jelmer Borst</i>	138
Learning to Cooperate by Anticipating Others' Mental States <i>Matteo Bortoletto</i>	139
The role of feedback in anticipatory human-computer-interaction <i>André Brechmann</i>	140
Anticipatory Human-Computer Interaction <i>Andreas Bulling</i>	141
AI-driven Personalization and Anticipatory HMI <i>Cristina Conati</i>	142
Cognitive Argumentation and Sensemaking <i>Emmanuelle Dietz</i>	144
Anticipation of User Performance For Adapting Visualization Systems <i>Steffen Frey</i>	145
Decision Making, Learning, and Adaptability in Human-Machine Collaborations <i>Cleotilde Gonzalez</i>	146
Integrating Models of Cognitive and Physical Human-Robot Interaction <i>Chenxu Hao</i>	147
Anticipatory Human-Machine Interaction to Explain AI Systems <i>Susanne Hindennach</i>	147
The Transportation Domain can Benefit From Anticipatory Models of Human-machine Interaction <i>Christian P. Janssen</i>	148
Probabilistic modelling for anticipating AI-assistants: Towards human-AI teams for experimental design and decision making <i>Samuel Kaski</i>	149
Surprise and Alignment: Their Role in Anticipation <i>Frank Keller</i>	150
Anticipatory Human-machine Interaction in Single-pilot Operations <i>Oliver W. Klaproth</i>	151

Proactive Machine Assistance in Visual Analytics <i>Alvitta Ottley</i>	151
Cognitive models of anticipating physical interactions <i>Nele Russwinkel</i>	152
Anticipatory Brain-Computer Interfaces <i>Andreea Ioana Sburlea</i>	153
Anticipation of User Goals from Human Behavioural Data for Human-AI Collaboration <i>Florian Strohm</i>	154
The Role of Skills in Anticipatory Human-Computer Interaction <i>Niels Taatgen</i>	154
Breakout Groups (Monday)	
Shared Representations and Communication <i>Alvitta Ottley, Matteo Bortoletto, Nele Russwinkel, Oliver Klaproth, Susanne Hindennach</i>	155
Timescales at which Anticipatory models for HMI can operate <i>Christian P. Janssen, Cristina Conati, Niels A. Taatgen</i>	156
Data-driven vs. rule-based models for anticipation <i>Jelmer Borst, Andreas Bulling, Emmanuelle Dietz, Chenxu Hao, Frank Keller, Florian Strohm, Andreea Ioana Sburlea, Samuel Kaski, Elisabeth André, Henny Admoni</i>	157
Breakout Groups (Tuesday and Wednesday)	
Formalisation of Anticipation and Use Cases <i>Cristina Conati, Emmanuelle Dietz, Steffen Frey, Susanne Hindennach, Alvitta Ottley, Niels A. Taatgen, Matteo Bortoletto, Henny Admoni</i>	160
Transportation-driven Anticipatory Human-Machine Interaction <i>Elisabeth André, Mihai Bâce, Jelmer Borst, André Brechmann, Andreas Bulling, Christian P. Janssen</i>	164
Planning in Anticipation <i>Leslie Blaha, Matteo Bortoletto, Samuel Kaski, Andreea Ioana Sburlea</i>	166
Physical Interaction <i>Nele Russwinkel, Frank Keller, Florian Strohm, Chenxu Hao, Henny Admoni</i>	167
Participants	169

3 Overview of Talks

3.1 Anticipation for Assistive and Collaborative Human-Machine Interaction

Henny Admoni (Carnegie Mellon University, USA, henny@cmu.edu)

License © Creative Commons BY 4.0 International license
© Henny Admoni

Joint work of Reuben Aronson, Abhijat Biswas, Maggie Collier, Kris Kitani, Ben Newman, Ada Taylor

Main reference Benjamin A. Newman, Reuben M. Aronson, Kris Kitani, Henny Admoni: “Helping People Through Space and Time: Assistance as a Perspective on Human-Robot Interaction”, *Frontiers Robotics AI*, Vol. 8, p. 720319, 2021.

URL <https://doi.org/10.3389/frobt.2021.720319>

Robots can help people live better lives by assisting on complex tasks involved in everyday activities. For example, robot arms can help people with motor impairments eat food or prepare a meal; robot servers can support waitstaff in a busy restaurant; and social robot sidekicks can improve the conversational agency of users with severe speech impairments. To be fully capable of providing fluent assistance, robots must understand *what* their human partner wants to do, *how* they would like to do it, and *when* the robot should step in.

This capability of providing *proactive assistance* relies on the robot’s understanding of implicit situational context, including their human partner’s intentions, preferences, and knowledge. A key insight that drives this research is that mental states, while not directly observable, are often linked to observable behaviors. For example, where someone gazes is tied to what task they’re trying to achieve, and hence their goal and likely future actions [1]. Thus, understanding the link between human behavior and mental states enables robots to anticipate human needs. Conversely, robots can use implicit signals as part of providing assistance to enable the human to anticipate the robot’s actions [2].

In this research, we develop algorithms that model assistance-relevant human mental states (e.g., needs, desires, and preferences) from observable behaviors, and produce robot assistance that is effective, fluent, and interpretable. The long-term research goal is to bridge the gap between human behavior and robot algorithms, which will enable researchers to develop more effective assistive and collaborative robots. In this multi-disciplinary work, we apply knowledge, tools, and methods from fields such as robotics, machine learning, artificial intelligence, computer perception, and cognitive science. We demonstrate the utility of proactive assistance across many levels of robot autonomy and in diverse domains, such as physical support on activities of daily living, driving support in autonomous vehicles, and social support during conversation

References

- 1 Aronson, Reuben M.; Admoni, Henny, *Gaze Complements Control Input for Goal Prediction During Assisted Teleoperation*, in *Proceedings of Robotics: Science and Systems*, 2022.
- 2 Taylor, Ada V.; Mamantov, Ellie; Admoni, Henny, *Wait Wait, Nonverbally Tell Me: Legibility for Use in Restaurant Navigation.*, in *Workshop on Social Robot Navigation at Robotics: Science and Systems*, 2021.

3.2 From Anticipatory to Socially-Aware User Interfaces

Elisabeth André (Augsburg University, DE, elisabeth.andre@uni-a.de)

License © Creative Commons BY 4.0 International license
© Elisabeth André

The development of anticipatory user interfaces is a key issue in human-centred computing. A major challenge of any anticipatory interface is to identify relevant information about the users including their environment that may be used by the system to prepare or adjust for future interactions.

In our very early work, we followed the paradigm of the so-called *anticipation feedback loop* [1] to improve the comprehensibility of multimodal presentation. The overall idea was to simulate how a user would perceive the output generated by a system and to adjust if necessary. For example, by parsing an utterance planned by the system, the generator made sure that the utterance did not contain unintended structural ambiguities. The implicit assumption was that we can predict the user's understanding processes from the system's understanding processes.

Our more recent work focuses on the paradigm of *Socially-Aware Interfaces* that adjust their behaviours based on an analysis of implicitly provided user feedback, such as social and affective signals, as an integral component of an anticipatory mechanism. For example, we investigated to what extent user engagement may be automatically detected from the user's gestures, postures, facial expressions and speech to adapt a robot's communicative style if necessary [2]. This approach is based on a direct observation of the user as opposed to a simulation. It bears the challenge that human expressions do not follow the precise mechanism of a machine, but are tainted with a high amount of variability, uncertainty and ambiguity.

References

- 1 Wahlster, W., André, E., Graf, W. & Rist, T. Designing Illustrated Texts: How Language Production Is Influenced By Graphics Generation, *5th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, pp. 8-14, 1991, <https://aclanthology.org/E91-1003/>
- 2 Ritschel, H., Baur, T. & André, E., Adapting a Robot's linguistic style based on socially-aware reinforcement learning, *26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* 2017, pp. 378-384, <https://doi.org/10.1109/ROMAN.2017.8172330>

3.3 Sensing and Modelling Human Attention towards Anticipatory Human-Computer Interaction

Mihai Băce (University of Stuttgart, DE, mihai.bace@vis.uni-stuttgart.de)

License © Creative Commons BY 4.0 International license
© Mihai Băce

The vision for the next generation of intelligent user interfaces (UIs) is to seamlessly assist users during everyday tasks, similar to how humans interact with one another. This is in stark contrast to current UIs, which are primarily reactive and lack the ability to understand or anticipate user actions, intentions, or goals. One fundamental requirement for developing systems that can anticipate human behaviour is the ability to sense and model human

attention. While human attention is key to many tasks in human-computer interaction (HCI), such as predicting interruptibility, boredom, or confusion, prior works have mostly modelled attention through proxies. Proxies such as user interactions or self-reports cannot fully capture the temporal, fine-grained dynamics of human attention.

One goal in my research is to leverage either sensed [1] or computational models of human attention for tasks such as predicting user interruptibility, modelling task interleaving, or in the context of work and productivity. Prior research has shown that, e.g. users can be in mental states that make them more likely to be interrupted [2]. My goal is to anticipate or predict when such mental states are likely to occur either using models of attention only, or in a multimodal setting (e.g. leveraging user interactions as well).

References

- 1 Bâce, M., Staal, S. & Bulling, A. Quantification of Users' Visual Attention During Everyday Mobile Device Interactions. *Proceedings Of The 2020 CHI Conference On Human Factors In Computing Systems*. (2020)
- 2 Mark, G., Iqbal, S., Czerwinski, M. & Johns, P. Focused, Aroused, but so Distractible: Temporal Perspectives on Multitasking and Communications. *Proceedings Of The 18th ACM Conference On Computer Supported Cooperative Work & Social Computing*. (2015)

3.4 Anticipatory Human-Machine Interaction with Cognitive Models

Jelmer Borst (University of Groningen, NL, j.p.borst@rug.nl)

License  Creative Commons BY 4.0 International license
© Jelmer Borst

Effective anticipatory human-machine interaction (aHMI) requires an accurate model of the human agent. To develop such models, cognitive architectures provide a promising starting point, as such architectures have long been used to develop detailed models of the human cognitive system at a level of abstraction that matches well to aHMI.

A direct example of such an approach is the Slimstampen system for fact learning ([3]; <http://www.slimstampen.nl>). The core of this system is a model of human memory taken from the cognitive architecture ACT-R (Anderson, 2007), which estimates the memory activation level of each fact that needs to be learned. Based on this information, the system anticipates the next fact that should be presented, thereby optimizing the scheduling of facts (e.g., present a fact again just before it will be forgotten). As a result, users can learn 10% more facts in the same amount of time as compared to standard methods such as flash-card learning.

As a second example, we used ACT-R to develop a model that explains under what circumstances interruptions are more or less disruptive [1]. Based on the detailed account of this model, we then built an interruption management system that measures pupil size to anticipate the best moments for interruptions [2]. By scheduling interruptions at opportune moments in the task, users finished the main task in 10% less time – while being interrupted just as often.


Together these examples show how the use of detailed cognitive models of the human mind can lead to very effective aHMI.

References

- 1 Borst, J., Taatgen, N. & Rijn, H. What makes interruptions disruptive? A process-model account of the effects of the problem state bottleneck on task interruption and resumption. *Proceedings Of The 33rd Annual ACM Conference On Human Factors In Computing Systems*. pp. 2971-2980 (2015)
- 2 Katidioti, I., Borst, J., Haan, D., Pepping, T., Vugt, M. & Taatgen, N. Interrupted by your pupil: An interruption management system based on pupil dilation. *International Journal Of Human-Computer Interaction*. **32**, 791-801 (2016)
- 3 Sense, F., Behrens, F., Meijer, R. & Rijn, H. An individual's rate of forgetting is stable over time but differs across materials. *Topics In Cognitive Science*. **8**, 305-321 (2016)

3.5 Learning to Cooperate by Anticipating Others' Mental States

Matteo Bortoletto (University of Stuttgart, DE, matteo.bortoletto@vis.uni-stuttgart.de)

License  Creative Commons BY 4.0 International license
© Matteo Bortoletto


Theory of Mind refers to humans' ability to infer and represent mental states of others. Such mental states include, for example, beliefs, desires and intentions. Computational Theory of Mind tries to build a model capable of inferring such states [1, 2]. As we want to infer mental states of other agents, Theory of Mind naturally deals with multi-agents systems, in which we have more than one agent observing or acting in the same environment. Multi-agents systems are quite difficult to deal with because one agent's action can influence other agents' actions. In this sense, anticipation plays a central role. By inferring beliefs, intentions and desires we can have cooperative agents capable of learning to assist humans or other agents [3]. Agents capable of anticipating others' behaviour can plan and act accordingly, avoiding mistakes or coordination errors. In principle, this strategy can be applied in both cooperative and competitive tasks or games. Multi-agent AI research has seen most success in competitive two-player zero-sum settings, like chess or Go. However, most real world interactions are cooperative, therefore applying Theory of Mind to improve human-AI cooperation, AI-AI cooperation or even human-human cooperation certainly represents one of the next big challenges in the field.

References

- 1 Rabinowitz, N., Perbet, F., Song, F., Zhang, C., Eslami, S. A., & Botvinick, M. (2018). Machine theory of mind. *International Conference on Machine Learning*, 4218–4227.
- 2 Chuang, Y.-S., Hung, H.-Y., Gamborino, E., Goh, J. O. S., Huang, T.-R., Chang, Y.-L., Yeh, S.-L., & Fu, L.-C. (2020). Using Machine Theory of Mind to Learn Agent Social Network Structures from Observed Interactive Behaviors with Targets. *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 1013–1019.
- 3 Bard, N., Foerster, J. N., Chandar, S., Burch, N., Lanctot, M., Song, H. F., Parisotto, E., Dumoulin, V., Moitra, S., Hughes, E., & others. (2020). The hanabi challenge: A new frontier for ai research. *Artificial Intelligence*, 280, 103216.

3.6 The role of feedback in anticipatory human-computer-interaction

*André Brechmann (Leibniz Institute for Neurobiology Magdeburg, DE,
brechmann@lin-magdeburg.de)*

License  Creative Commons BY 4.0 International license
© André Brechmann

In human-computer-interaction the action of the technical system can be viewed as feedback to the prior action of the user. Such feedback can fulfil a number of different functions, e.g. signalling that a user's action has been registered by the technical system, execution of a predefined function, actively provide task-related information, providing motivational feedback, or evaluate a user's decision. In the neurosciences feedback is inextricably linked with reward rooted in traditional reinforcement-based theories that were developed based on animal studies that applied reward and/or punishment for the animal to learn. This research has even coined the term reward system of the brain consisting of midbrain dopaminergic areas (substantia nigra/VTA) and main projection areas, e.g. the striatum and prefrontal cortex. Following this tradition, the vast majority of studies on the neural correlates of feedback processing in humans have been designed as reinforcement experiments using (monetary) rewards. Consequently, any findings of activation of the brain's "reward" system has almost exclusively been attributed to mechanisms of reward processing (see Wolff et al., 2020 for references). When we entered the human-computer-interaction domain, we started with the most basic function of feedback, i.e. to signal that a users decision (button press) has been registered by the technical system to study effects of delayed and omitted system response and adaptive processes in the user's brain towards delayed system responses (Kohrs et al. 2012, 2016). To our surprise, we showed the reward system of the brain to be specifically involved in the processing of such registering feedback. A consistent interpretation of these findings relates to the user making predictions about the feedback of the technical system independent of any reward. For me, this opened up a novel neuroscience perspective on feedback processing in human-computer-interaction because humans anticipate reactions of interactions partners and probably assume such skills to be present in intelligent technical systems, too. We are currently interested in understanding how a technical system can support a user in the course of learning. As use case, we designed a multidimensional category learning paradigm where humans learn the conjunction of two rules by trial-and-error within a session of 180 trials. We observed that many participants persistently stick to a strategy that only considers one to the two rules. In order to prevent the consolidation of such an incorrect strategy, we implemented a heuristic that anticipates such behaviour based on the decision history of the learners. This heuristic then triggers a feedback intervention providing helpful information towards the solution of the learning problem. The results of this intervention in terms of subsequent performance and subjective ratings by the learner how helpful they perceived the intervention was a key motivation to pursue the topic of anticipatory human-computer-interaction: Even though the intervention was objectively helpful, many of the subjects did not rate the support very highly even if they improved in performance. We are now working towards understanding the reasons for this adverse effects of feedback intervention. A simple explanation could be that learners did not like to be interrupted at all or at a later moment or that the mode of neural processing was not compatible with the need to reflect on the previous performance. Whatsoever, my future goal is to include the patterns of brain activity during the course of learning into the process of anticipating the need and the best moment for supportive feedback intervention and to learn from the impact of an intervention on behavioural, psycho- and neuro-physiological data for anticipating the consequences of a proactive intervention of a technical system.

References

- 1 Kohrs C, Angenstein N, Scheich H, Brechmann A. 2012. Human striatum is differentially activated by delayed, omitted, and immediate registering feedback. *Frontiers in Human Neuroscience*. 6(AUGUST). <https://doi.org/10.3389/fnhum.2012.00243>
- 2 Wolff S, Kohrs C, Angenstein N, Brechmann A. 2020. Dorsal posterior cingulate cortex encodes the informational value of feedback in human-computer interaction. *Scientific Reports*. 10(1):Article 13030. <https://doi.org/10.1038/s41598-020-68300-y>

3.7 Anticipatory Human-Computer Interaction

Andreas Bulling (*University of Stuttgart, DE, andreas.bulling@vis.uni-stuttgart.de*)

License © Creative Commons BY 4.0 International license
© Andreas Bulling

Main reference Julian Steil, Philipp Müller, Yusuke Sugano, Andreas Bulling: “Forecasting user attention during everyday mobile interactions using device-integrated and wearable sensors”, in Proc. of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services, MobileHCI 2018, Barcelona, Spain, September 03-06, 2018, pp. 1:1–1:13, ACM, 2018.

URL <https://doi.org/10.1145/3229434.3229439>

Even after three decades of research on human-computer interaction (HCI), current general-purpose user interfaces (UI) still lack the ability to attribute mental states to their users, i.e. they fail to understand users’ intentions and needs and to anticipate their actions. This drastically restricts their interactive capabilities. We aim to establish the scientific foundations for a new generation of user interfaces that pro-actively adapt to users’ future input actions by monitoring their attention and predicting their interaction intentions – thereby significantly improving the naturalness, efficiency, and user experience of the interactions. Realising this vision of *anticipatory human-computer interaction* requires groundbreaking advances in everyday sensing of user attention from eye and brain activity. We further require methods to predict entangled user intentions and forecast interactive behaviour with fine temporal granularity during interactions in everyday stationary and mobile settings. Finally, we require fundamental interaction paradigms that enable anticipatory UIs to pro-actively adapt to users’ attention and intentions in a mindful way. We have identified four challenging application areas where these new capabilities have particular relevance: 1) mobile information retrieval, 2) intelligent notification management, 3) Autism diagnosis and monitoring, and 4) computer-based training. Anticipatory human-computer interaction offers a strong complement to existing UI paradigms that only react to user input post-hoc. If successful, work on this topic will deliver the first important building blocks for implementing Theory of Mind in general-purpose UIs. As such, anticipatory HCI has the potential to drastically improve the billions of interactions we perform with computers every day, to trigger a wide range of follow-up research in HCI as well as adjacent areas within and outside computer science, and to act as a key technical enabler for new applications, e.g. in healthcare and education.

References

- 1 Müller, P., Sood, E. & Bulling, A. Anticipating Averted Gaze in Dyadic Interactions. *Proc. ACM International Symposium On Eye Tracking Research And Applications (ETRA)*. pp. 1-10 (2020)
- 2 Hu, Z., Bulling, A., Li, S. & Wang, G. FixationNet: Forecasting Eye Fixations in Task-Oriented Virtual Environments. *IEEE Transactions On Visualization And Computer Graphics (TVCG)*. 27, 2681-2690 (2021)

- 3 Strauch, C., Hirzle, T., Stigchel, S. & Bulling, A. Decoding binary decisions under differential target probabilities from pupil dilation: A random forest approach. *Journal Of Vision (JOV)*. 21, 1-13 (2021)

3.8 AI-driven Personalization and Anticipatory HMI

Cristina Conati (Department of Computer Science, University of British Columbia, CA, conati@cs.ubc.ca)

License  Creative Commons BY 4.0 International license
© Cristina Conati

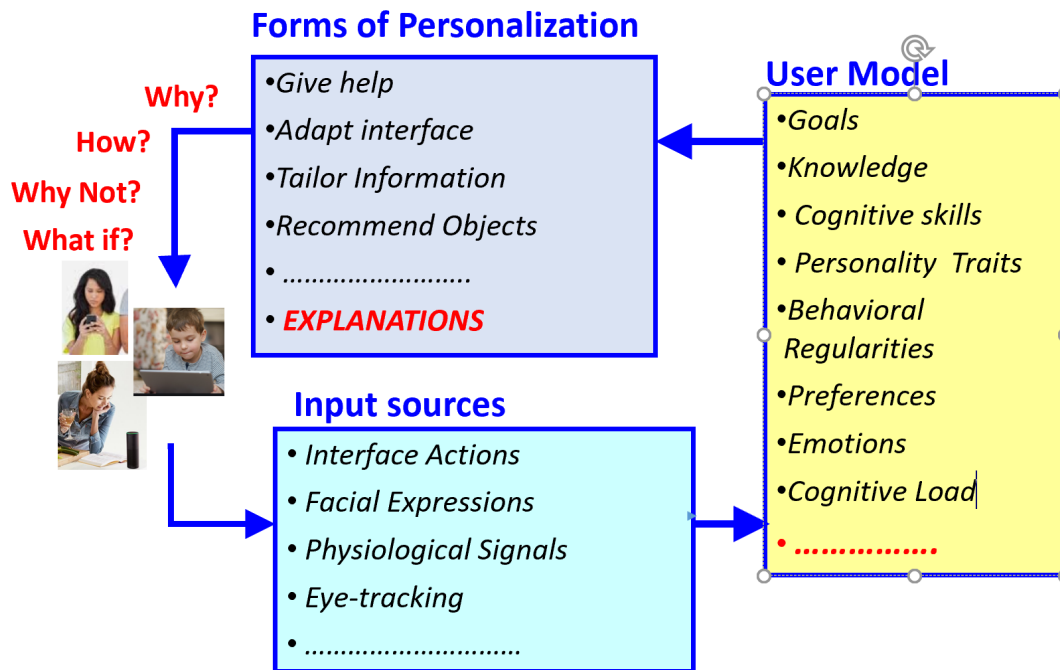
My research focuses on AI-Driven personalization, an interdisciplinary field at the intersection of Artificial Intelligence (AI), Human Computer Interaction (HCI) and Cognitive Science. The general objective of this field is to support effective AI-human collaboration by enabling AI systems to personalise their actions and behaviours to the specific needs, states, preferences and abilities of their individual user. Examples of active research in this field include recommender systems, intelligent-tutoring systems, conversational agents, and affect-aware systems.

To provide personalised interaction, an AI system needs to have a *personalization loop* in which it acquires a model of its user by inferring relevant user properties (see User Model in Figure 1, right) from available interaction behaviors (see Input Sources in Figure 1, bottom) and decides how to personalise the interaction accordingly (see Forms of Personalization in Figure 1, top). In my research I have been investigating AI-driven personalization in the context of Intelligent Tutoring Systems (ITS), intelligent help agents for complex interfaces, and personalized support to visualization processing.

My work in ITS revolves around the objective of moving beyond systems designed to provide support for problem solving, by modeling and adapting to students meta-cognitive abilities and by targeting educational activities that are more open-ended in nature, such as learning from interactive simulations or playing educational games, e.g., [2, 3]. I have done extensive work on using eye-tracking data to inform user modelling, showing that information from eye-tracking can help predict both user *short-term states* (e.g. confusion, affective valence, learning) as well as *long-term* cognitive abilities (e.g. perceptual skills, reading proficiency) and traits (e.g. personality). A key application of eye-tracking for user modelling is in driving personalization for tasks that are mainly perceptual in nature, such as processing data visualisations, which has been the focus of my research on user-adaptive visualisations [1]. The key idea underlying this research is to complement traditional approaches to visualisation design based on properties of the target data and tasks by providing real-time support to visualisation processing driven by relevant users states (e.g. attention, confusion) and abilities (e.g. levels of visual literacy, perceptual speed, visual memory etc.).

My most recent research proposes the concept of personalised Explainable AI. This work extends the recent interest in AI systems that can explain their inner workings to their end-users with the idea that these explanations should be part of the personalisation loop (see Figure 1, top), namely they should be dynamically personalised to factors including context, task criticality and user differences (e.g., cognitive abilities and transient states like confusion or cognitive load).

A distinction that needs to be made to understand how AI-driven personalization relates to Anticipatory Human-Machine Interaction (AHMI) is what triggers personalization (often referred to as adaptation in the literature), namely whether the AI agent is personalising



■ **Figure 1** Adaptive Loop for AI-Driven personalization, with explanations of the AI actions added as a form of personalization.

to a *current* user state that it has detected/assessed/perceived or to a user state that is *predicted* to be happening at some point in the future. When the personalization is driven by the current user state, it is *reactive* in nature¹. Personalization driven by a *predicted* future state of the user can be viewed as *preventive/proactive/anticipatory*. I see the latter as being the key aspect of AHMI. I will illustrate this distinction with examples based on some of my work.

An example of *reactive personalization* in my ITS research comes from my work on ITS for supporting students in acquiring meta-cognitive skills relevant for learning. Specifically, in [2] we focused on the meta-cognitive skill known as self-explanation, i.e. one’s tendency to explain and elaborate instructional material to themselves when studying. There is ample evidence in Cognitive Psychology that this skill is conducive to better learning, and that students who lack the skill can acquire it if they undergo instructional sessions during which a tutor prompts them to self-explain if they do not do it. During these sessions, students are asked to self-explain aloud, just so that the tutor can detect if they are self-explaining or not. However, self-explanation usually happens in one’s head. In [2], we worked on an ITS that assesses if a student self-explains in their head while studying given material, based on their gaze patterns and interface actions. The ITS then reacts if it detects that self-explanation did not happen for a specific aspect of the instructional material, by generating prompts to encourage and guide the student to self-explain that material.

An example of *anticipatory personalization* in my ITS research comes from my work on ITS for supporting students learning through explorations in open-ended-learning environments.

¹ Note that this reactive personalization is different, and more sophisticated, than the reactivity embedded in traditional interfaces, where each interface action (e.g. a button press) is associated by design to a specific system’s response, which does not depend on the user who performed the action.

The capability to explore effectively is relevant to many tasks involving interactive systems, but not all users possess this ability. I have been investigating ITS that can provide personalized real-time support to users who are having difficulties with an exploratory task, while interfering as little as possible with the unconstrained nature of the interaction. Building such ITS is challenging because it requires having a user model that can predict the effectiveness of open-ended behaviours for which there is often no formal definition of correctness. Our approach relies on clustering and association rule mining to discover from data the user exploratory behaviours conducive or detrimental to learning. Supervised machine learning is applied to the resulting clusters to derive classifiers that predict in real-time whether a user will learn from the interaction and, if not, it anticipates how the user behavior should change to increase their learning. Based on this anticipation, the ITS then generates hints designed to foster the desired change in behavior [3].

The distinction between reactive and anticipatory personalization is one that has not yet been made in the literature, but it is important to think about the unique advantages and challenges that anticipatory personalization brings to AI-Human collaboration.

References

- 1 Conati, C., Carenini, G., Toker, Derek., Lallé, S., *Towards User-Adaptive Information Visualization*, Twenty-Ninth AAAI Conference on Artificial Intelligence, 2015.
- 2 Conati, C., Merten, C., *Eye-tracking for user modeling in exploratory learning environments: An empirical evaluation*, Knowledge Based Systems 20(6): 557-574, 2007.
- 3 Kardan, S., Conati, C., *Providing Adaptive Support in an Interactive Simulation for Learning: An Experimental Evaluation*, Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, 2015.

3.9 Cognitive Argumentation and Sensemaking

Emmanuelle Dietz (Airbus Central R&T – Hamburg, DE, emmanuelle.dietz@airbus.com)

License  Creative Commons BY 4.0 International license
 Emmanuelle Dietz

Systems built with the purpose of human-machine interaction do not necessarily need to behave optimally in each circumstance when interacting with humans. However, they should have the ability to provide socially acceptable explanations for their decisions. They might generate such explanations by having an internal representation of the current state of the human mind (theory of mind). This representation might allow the system to predict future events and possibly prepare by intervening with some action (e.g. to avoid catastrophic outcomes or to improve the human's situation).

The internal representation of predicting and deciding on which action to choose can happen on various levels. An important requirement however, is that the system should have the ability to provide socially acceptable explanations for their decisions. More specifically, their explanations should have some basic reasons of support (attributive), explain why a conclusion is supported in contrast to opposing conclusions (contrastive) and provide information that guides on how to act following the conclusion (actionable). Cognitive argumentation fulfills these requirements. In cognitive argumentation, having its roots in computational argumentation, the dialectic argumentation process is assumed to be heavily guided by biases or heuristics [1, 2]. Initially, a root argument for a certain position is built (e.g. based on the awareness in the current situation), (strong) counterarguments are

searched (e.g. factual information contracts the argument), which the root argument defends, and is either replaced by stronger arguments or extended with the defense. This process seems similar to sensemaking, in which humans generate mental models to explain events, question these models, and reframe or replace these models by more adequate ones.

In the context of human-machine interaction anticipatory thinking can be implemented through future oriented cognitive argumentation by generating future arguments for or against carrying out a possible future action (of the human or machine) that leads to the most desired state.

References

- 1 Dietz, E.; Kakas, A., *Cognitive Argumentation and the Selection Task*. Proceedings of the Annual Meeting of the Cognitive Science Society, pp. 1588–1594, 2021.
- 2 Dietz, E., *Argumentation-based Reasoning guided by Chunk Activation in ACT-R*. In Proceedings of the 20th International Conference on Cognitive Modelling, 2022.

3.10 Anticipation of User Performance For Adapting Visualization Systems

Steffen Frey (University of Groningen, NL, s.d.frey@rug.nl)

License  Creative Commons BY 4.0 International license
© Steffen Frey

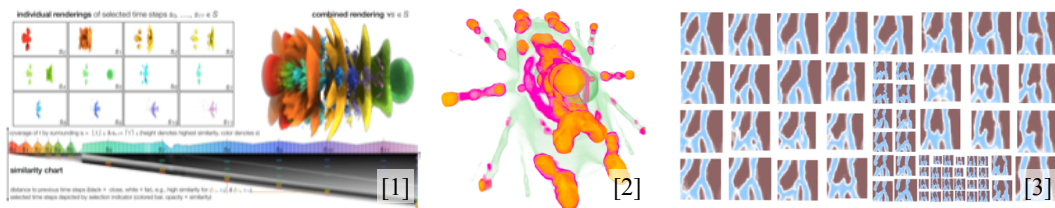
My research concerns the development of visualization methods to gain insights from large quantities of scientific data (typically acquired from extensive experiments and simulations). Interactive approaches enable explorative user analysis of rich and complex data, but respective visualization systems generally exhibit numerous and degrees of freedom regarding how information is presented and explored.

In many of my previous works, techniques and systems are configured in a data-driven way, e.g., which time steps [1] or spatio-temporal subsets [2] to convey and how to layout large numbers of graphical tiles depicting members of an ensemble [3] (Figure 2). For this, these approaches generally optimize application-specific metrics or employ self-supervised learning. For example, metrics quantify how well dynamics in a full time series are captured by a subset [1] or the similarity of ensemble members positioned nearby in a grid layout [3]), and learning can be used identify unusual behavior [2].

Conceptually, these approaches optimize the transformation of the data into an (interactive) visualization, but largely neglect – or only indirectly consider – the human decision-making process operating on this basis. Explicitly incorporating this aspect, the anticipation of user performance for a given (configuration of) visualization system allows to explicitly optimize the visual representation and interface for a user to gain insights more quickly and/or more accurately.

References

- 1 Frey, S.; Ertl, T., *Flow-based Temporal Selection for Interactive Volume Visualization*, Computer Graphics Forum, 2017.
- 2 Tkachev, G., Frey, S., Ertl, T., *Local Prediction Models for Spatiotemporal Volume Visualization*, IEEE Transactions on Visualization and Computer Graphics, 2021.
- 3 Frey, S., *Optimizing Grid Layouts for Level-of-Detail Exploration of Large Data Collections*, Computer Graphics Forum, 2022.



■ **Figure 2** Examples of data-driven visualization via the optimization of metrics [1,3] and unsupervised learning [2].

3.11 Decision Making, Learning, and Adaptability in Human-Machine Collaborations

Cleotilde Gonzalez (Carnegie Mellon University, USA, coty@cmu.edu)

License © Creative Commons BY 4.0 International license
© Cleotilde Gonzalez

A major challenge for research in artificial intelligence is to develop systems that can infer the goals, beliefs, and intentions of others (i.e., systems that have theory of mind, ToM) (Nguyen & Gonzalez, 2021). To generate relevant Human-Machine collaborations, one needs to create computational representations of human behavior. In the past decades cognitive models have advanced towards more accurately representing and replicating human cognitive processes. In particular, we have been involved in generating models that can explain how humans make decisions, learn, and adapt to improve their choices in dynamic tasks. We rely on Instance-Based Learning Theory (Gonzalez, Lerch & Lebiere, 2003; Nguyen, Phan, & Gonzalez, 2022) to construct models of human dynamic decisions in a large number of situations. Our applications for anticipatory Human-Machine Collaborations include but are not limited to: Applications in cybersecurity, replicating attacker’s behaviors that can be used to inform defenders; and replicating end-users’ phishing behaviors that can help prevent successful attacks. We have also used these models in dynamic control tasks and scenarios for search and rescue tasks, where predicting the state of the mind of teammates’ decisions is essential.

References

- 1 Gonzalez, C.; Lerch, F.J.; Lebiere, C., *Instance-based learning in dynamic decision making*. Cognitive Science 27, pp. 591–635, 2003.
- 2 Nguyen, T.N.; Gonzalez, C., *Theory of Mind from Observation in Cognitive Models and Humans*. Topics in Cognitive Science, 2021.
- 3 Nguyen, T.N.; Phan, N.D.; Gonzalez, C., *SpeedyIBL: A Comprehensive, Precise, and Fast Implementation of Instance-Based Learning Theory*. Behavior Research Methods, 2022.

3.12 Integrating Models of Cognitive and Physical Human-Robot Interaction

Chenxu Hao (Friedrich-Alexander-Universität Erlangen-Nürnberg, DE, chenxu.hao@fau.de)

License  Creative Commons BY 4.0 International license
© Chenxu Hao

Joint work of Chenxu Hao, Nele Russwinkel, Daniel F.B. Haeufle, Philipp Beckerle

Research in human-robot interactions (HRI) usually focuses on either cognitive interactions or physical interactions. However, characteristics of HRI in real-life scenarios often involve both information exchanges on the cognitive level and force exchanges on the physical level. Specifically, the human agent and the robotic agent need to infer each other's intentions, predict each other's actions, and act adaptively in the given environment [1]. In such dynamic interactions, the robotic agent also needs to anticipate in order to support the human agent with flexibility. Therefore, it is crucial to create unified models of HRI that can capture the dynamic exchanges on both cognitive and physical levels while taking the environment into account (see [2], for an example of such a unified model of a single agent). Our work aims to provide a conceptual framework with possibilities to connect models of physical and cognitive HRI. We also hope to potentially apply our framework to an anticipatory robotic agent that provides the human agent with different levels of assistance.

References

- 1 Ho, M. K., & Griffiths, T. L. *Cognitive science as a source of forward and inverse models of human decisions for robotics and control*. arXiv:2109.00127 [cs, eess], 2021.
- 2 Kahl, S., Wiese, S., Russwinkel, N., & Kopp, S. *Towards autonomous artificial agents with an active self: modeling sense of control in situated action*. *Cognitive Systems Research*, 72, 50–62, 2022.

3.13 Anticipatory Human-Machine Interaction to Explain AI Systems

Susanne Hindennach (University of Stuttgart, DE, susanne.hindennach@vis.uni-stuttgart.de)

License  Creative Commons BY 4.0 International license
© Susanne Hindennach

Humans have the astounding ability to attribute mental states to their interaction partners and reason about their behaviour – so-called Theory of Mind (ToM). Similarly, humans also attribute a mind and human abilities to AI systems (like intelligence and learning). The attribution of a human-like mind to the AI system puts the focus on the AI systems, thereby losing sight of the human decisions about who and what gets counted when building them. The thought processes, intents and decisions can be surfaced by redirecting Theory of Mind away from the AI system and towards the AI experts who built the system. The redirection necessitate anticipatory human-machine interaction in order to (1) infer the mental states of the AI experts and (2) predict the explanations needed by the user of the AI system. In previous work, we have shown that intents – a mental state with explanatory power – can be predicted from mouse and keyboard action in a text formatting task[1].

References

- 1 Zhang, G., Hindennach, S., Leusmann, J., Bühler, F., Steuerlein, B., Mayer, S., Bâce, M. & Bulling, A. Predicting Next Actions and Latent Intents during Text Formatting. *Proceedings Of The CHI Workshop Computational Approaches For Understanding, Generating, And Adapting User Interfaces*. pp. 1-6 (2022)

3.14 The Transportation Domain can Benefit From Anticipatory Models of Human-machine Interaction

Christian P. Janssen (Utrecht University, NL, c.p.janssen@uu.nl)

License © Creative Commons BY 4.0 International license
© Christian P. Janssen

Main reference Christian P. Janssen, Shamsi T. Iqbal, Andrew L. Kun, Stella F. Donker: “Interrupted by my car? Implications of interruption and interleaving research for automated vehicles”, *Int. J. Hum. Comput. Stud.*, Vol. 130, pp. 221–233, 2019.

URL <https://doi.org/10.1016/j.ijhcs.2019.07.004>

At this Dagstuhl seminar, I have taken the position that the transportation domain can benefit from anticipatory models for human-machine interaction. One specific domain that I have investigated over the last few years is semi-automated driving. Here, the vehicle takes over some components of the driving task (such as lateral or longitudinal control) under specific conditions (or operational design domains; such as highway roads). However, typically, the vehicle cannot drive under all circumstances, and some human intervention is occasionally needed. The default model in the literature is to look at what the minimum time is that humans need between a warning and the time where they need to act. There seems to be consensus that somewhere between 5 to 8 seconds is a minimum time, and indeed the large majority of the literature has looked at response time situations of around 2 seconds. However, as I have argued elsewhere in more detail [2], although it is beneficial for research to understand what the minimum response interval needs to be, there can be many situations where it is useful to think more ahead and maybe have “pre-alerts” or forewarnings. One reason might be that as automated vehicles gain more capacity, human drivers might not always pay as much attention to the as they would under non-automated conditions, and start performing other tasks. Working on non-driving related activity is something that some drivers already do under regular (non-automated) driving, but that they want to do under higher levels of automation. In these situations, where attention has been to another task or activity than the driving task, one cannot assume that the attention of the driver has been on the road and that they had sufficient situation awareness to act in future (emergency) situations. Moreover, they might not even process the alert itself instantly. So, instead of thinking about fast task switches within, say, 5-8 seconds, one can instead think of how the process of returning attention from a non-driving activity to driving unfolds. I have proposed a conceptual model to capture this [2], of which the first empirical tests have taken place [2]. Now, the next step is to develop models that try to analyze human state and act accordingly, to interrupt at opportune moments. Such models need to take into account the user state, but also the environment and the driving scenario (e.g., is it a situation where there are many alternative actions, or is the intended action quickly clear from the context?). Personally, I would hope that techniques such as Hidden Markov Models can be effective in capturing human states and its dynamics [3]. Without anticipatory models, interruptions and alerts might come at less opportune moments, which can lead to increased stress or workload, and leave the driver in no capacity to act quickly and safely. Although I have specified these principles for driving, I anticipate that they also apply to other traffic domain such as railroads, ships, and airplanes.


References

- 1 C. P. Janssen, S. T. Iqbal, A. L. Kun, and S. F. Donker, “Interrupted by my car? Implications of interruption and interleaving research for automated vehicles,” *International Journal of Human Computer Studies*, vol. 130, 2019, doi: 10.1016/j.ijhcs.2019.07.004.

- 2 D. Nagaraju, A. Ansah, A., N.A.N. Ch, C. Mills, C.P. Janssen, O. Shaer, & A.L. Kun, “How Will Drivers Take Back Control in Automated Vehicles? A Driving Simulator Test of an Interleaving Framework,” in 13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, 2021, pp. 20–27. doi: 10.1145/3409118.3475128.
- 3 C. P. Janssen, L. N. Boyle, A. L. Kun, W. Ju, and L. L. Chuang, “A Hidden Markov Framework to Capture Human–Machine Interaction in Automated Vehicles,” *International Journal of Human-Computer Interaction*, vol. 35, no. 11, pp. 947 – 955, 2019, doi: 10.1080/10447318.2018.1561789.

3.15 Probabilistic modelling for anticipating AI-assistants: Towards human-AI teams for experimental design and decision making

Samuel Kaski (Aalto University, FI & University of Manchester, UK, samuel.kaski@aalto.fi)

License  Creative Commons BY 4.0 International license
© Samuel Kaski

Current AI-assistants are only able to help us when we can precisely specify what we want – that is, the goal and the problem setting. However, that arguably is not when we most need assistance, and effectively the current assistants require us to program them, just with a different kind of an interface such as giving precise spoken commands in restricted language. Instead, we need assistance the most we do not *yet* know precisely what we want, because the goal may be tacit, unclear or evolving. I am developing AI-assistance principles for situations such as drug design or scientific research, where the goal initially is unclear and evolves during the process.

Such assistants need to anticipate their users to be helpful to them: both their actions, some of which they can assume to arise from the user’s planning towards their goals, but also their goals and the development of the goals, and furthermore how the users would understand the assistant’s actions. This is particularly difficult because of the nature of the design-type tasks, where each task is different and laborious, resulting in scarcity of data for learning the goals. In particular, this culminates in “zero-shot assistance” for design and decision making, which I believe requires (i) probabilistic modelling to handle the uncertainty arising from scarcity of data and a large number of potential explanations of behaviour, (ii) nested POMDP-type modelling to capture a “theory of mind” of the user, to both infer the goal by reverse-engineering the behaviour and to anticipate how the user would understand the assistant’s action, and (iii) novel computational approximations to actually do the necessary inferences.

As a result, we hope to have better tools for solving the grand challenge problems humanity currently is facing.

References

- 1 Mustafa Mert Çelikok, Frans A. Oliehoek, and Samuel Kaski. Best-response Bayesian reinforcement learning with BA-POMDPs for centaurs. In *Proc. AAMAS 2022*, accepted for publication.
- 2 Sebastiaan De Peuter and Samuel Kaski. Zero-Shot Assistance in Novel Decision Problems. arXiv:2202.07364, 2022.
- 3 Antti Kangasrääsiö, Jussi P.P. Jokinen, Antti Oulasvirta, Andrew Howes, and Samuel Kaski. Parameter inference for computational cognitive models with approximate Bayesian computation. *Cognitive Science*, 43:e12738, 2019.

3.16 Surprise and Alignment: Their Role in Anticipation

Frank Keller (*University of Edinburgh, UK, keller@inf.ed.ac.uk*)

License © Creative Commons BY 4.0 International license
© Frank Keller

Main reference Moreno I. Coco, Rick Dale, Frank Keller: “Performance in a Collaborative Search Task: The Role of Feedback and Alignment”, *Top. Cogn. Sci.*, Vol. 10(1), pp. 55–79, 2018.

URL <https://doi.org/10.1111/tops.12300>

In psycholinguistics, surprise is a key concept. It measures how predictable the next word in a sentence is, given the prior context of the word. Surprise has been shown to correlate with behavioral measures such as cloze probabilities, reading times [1], or EEG signatures such as the N400. Mathematically, surprise is the negative log of the conditions probability of word n given words $1 \dots n$, where a language model is typically used to estimate this probability. Recent work in computational cognitive modeling has used surprise as a component in a more comprehensive cognitive model. For example, the Neural Attention Tradeoff model of [2] models the allocation of human attending during reading: it predicts which words are skipped by the reader, and which ones are fixated. Skipping in the model depends on surprise, but also on aspects of the reading task (e.g., reading for pleasure, proofreading, reading to answer a question).

There is an intuitive connection between anticipation and surprise. Assume an agent predicts the next word, and based on that decides whether to skip it or not. If that decision turns out to be incorrect, then the agent will be surprised, which is something we can measure behaviorally. Surprise can therefore be conceptualized as unsuccessful anticipation. It is worth exploring this idea in the present seminar, and in particular discuss to what extent surprise generalizes to domains other than language processing.

Cognitive scientists have not only studied language processing in single readers and speakers, but have also investigated pairs of speakers (dyads) that communicate and interact to solve a shared task. In such a situation, dialogue participants are known to align with each other. This means that their linguistic productions become more similar across a whole range of dimensions: they use similar words, similar syntactic structures, adapt their speech rate and even their pronunciation. At the same time, they also align non-verbal behavior such as gaze, posture, and gesture. The degree of alignment typically increases over the course of the dialogue, and alignment is correlated with task success. A classical conceptualization of this process in the Interactive Alignment Model [3]. More recent work has refined our understanding of alignment: for instance, using a collaborate search task, Coco et al. (2018) show that the relationship between alignment and task success is not absolute, but depends on how interactive the task is.

Alignment is relevant in the context of anticipation. A participant that aligns with their collaborator predicts the future behavior of that collaborator and decides to make their own behavior more similar. Alignment is therefore a form of long-term anticipation and, like anticipation more generally, it makes collaborating on a task more efficient. This is another observation that could feed into the discussion at this seminar, where questions to be explored include: Does the relationship between alignment and anticipation generalize to non-verbal tasks? Does alignment also make human-machine interaction more efficient? Should machines align with humans? Are humans able to align with machine?

References

- 1 Demberg, Vera and Frank Keller. 2008. Data from Eye-tracking Corpora as Evidence for Theories of Syntactic Processing Complexity. *Cognition* 109:2, 193–210.

- 2 Hahn, Michael, and Frank Keller. 2018. Modeling Task Effects in Human Reading with Neural Attention. <https://arxiv.org/abs/1808.00054>
- 3 Pickering, Martin J. and Simon Garrod. 2004. *Toward a mechanistic psychology of dialogue*. Behavioral and Brain Sciences 27:2, 169–190.

3.17 Anticipatory Human-machine Interaction in Single-pilot Operations

Oliver W. Klaproth (Airbus – Hamburg DE, oliver.klaproth@airbus.com)

License © Creative Commons BY 4.0 International license
© Oliver W. Klaproth

Main reference Sebastian Blum, Oliver Klaproth, Nele Russwinkel: “Cognitive Modeling of Anticipation: Unsupervised Learning and Symbolic Modeling of Pilots’ Mental Representations”, Topics in Cognitive Science, Vol. 14, pp. 718-738, 2022.

URL <https://doi.org/10.1111/tops.12594>

The interaction of two pilots in the cockpit benefits from their ability to anticipate each other’s behaviour. Single pilot operations will be characterised by increased interaction with higher levels of automation instead of with another human pilot. By modelling the pilot and their mental representation of the automation, the automation can anticipate the pilot’s behaviour and adapt its own behaviour accordingly to stay in line with the pilot’s anticipations [1]. Cognitive architectures like ACT-R have been used to model pilots [2, 3]. In particular, instance-based learning in ACT-R has been shown to be effective in learning mental representations through observations of behaviour sequences and in modelling intuitive associative processes underlying anticipation. We hypothesise that instance-based ACT-R models of pilots, for example based on training on flight recorder data, can enable anticipatory human machine interaction in the cockpit during single pilot operations.

References

- 1 Blum, S., Klaproth, O. & Russwinkel, N. Cognitive Modeling of Anticipation: Unsupervised Learning and Symbolic Modeling of Pilots’ Mental Representations. *Topics In Cognitive Science*. (2022)
- 2 Klaproth, O., Halbrügge, M., Krol, L., Vernaleken, C., Zander, T. & Russwinkel, N. A neuroadaptive cognitive model for dealing with uncertainty in tracing pilots’ cognitive state. *Topics In Cognitive Science*. **12**, 1012-1029 (2020)
- 3 Klaproth, O., Vernaleken, C., Krol, L., Halbruegge, M., Zander, T. & Russwinkel, N. Tracing pilots’ situation assessment by neuroadaptive cognitive modeling. *Frontiers In Neuroscience*. **14** pp. 795 (2020)

3.18 Proactive Machine Assistance in Visual Analytics

Alvitta Ottley (Washington University in St. Louis – St. Louis, USA, alvitta@wustl.edu)

License © Creative Commons BY 4.0 International license
© Alvitta Ottley

The visual analytics community has made significant strides in developing systems that facilitate the interplay between humans and machines in exploratory data analysis and sensemaking [1]. These advances have enabled us to learn from user interactions and uncover their analytic goals (e.g., [2, 3]). Moreover, they have set up the foundation for creating visual analytic systems that anticipate and assist or guide users during data exploration. Providing

such guidance will likely become more critical as datasets grow in size and complexity, overwhelming the limited screen real estate and human cognitive power. My team and I use interdisciplinary approaches to identify optimal data representations for effective decision-making and design human-in-the-loop visual analytics interfaces more attuned to domain scientists' exploration goals. Still, some open anticipatory challenges and considerations may be unique to the visual analytics setting. In particular, the role of the AI may depend on the task and include but is not limited to: (1) the AI can be a teammate, working with the human to collaboratively solve a problem by either performing the same task concurrently or separate complementary sub-tasks. (2) the AI may be an assistant, observing actions and taking preparatory steps such as prefetching and precomputation. (3) The AI may serve as a learner, proactively observing actions to create a knowledge graph or process script. Moreover, the user's tasks will likely evolve. Thus, anticipation should happen at a global level (e.g., deciding the role the AI needs to assume based on the observed state) and at a local level (e.g., deciding which data points to prefetch). Finally, the community should address ethical considerations for how AI interventions might influence the user's analysis. For example, we need to preemptively consider how to track biases in the data shown to the user.

References

- 1 Crouser, R., Ottley, A. & Chang, R. Balancing human and machine contributions in human computation systems. *Handbook Of Human Computation*. pp. 615-623 (2013)
- 2 Ottley, A., Garnett, R. & Wan, R. Follow the clicks: Learning and anticipating mouse interactions during exploratory data analysis. *Computer Graphics Forum*. **38**, 41-52 (2019)
- 3 Monadjemi, S., Garnett, R. & Ottley, A. Competing models: Inferring exploration patterns and information relevance via bayesian model selection. *IEEE Transactions On Visualization And Computer Graphics*. **27**, 412-421 (2020)

3.19 Cognitive models of anticipating physical interactions

Nele Russwinkel (Technische Universität Berlin, DE, nele.russwinkel@tu-berlin.de)

License  Creative Commons BY 4.0 International license
© Nele Russwinkel

For gaining mutual understanding between a human and an intelligent agent the intelligent agent needs to develop a model of the human partner and the environment as well as of the task. Depending on the type of interaction and purpose of the interaction different requirements arise for a fluent and efficient collaboration. For more complex interactions the assumption of a shared representation about the individual goals, the task and situation understanding is essential. The assumed shared representation of technical system and human might differ. Direct and indirect communication is necessary to provide information about the individual shared representations. Especially for physical Interactions with technical Interactions a lot of different cues make sure that e.g. a handing over procedure of an object will be well coordinated in time, location and type. What types of cognitive models are needed to address these issues? Since anticipation is the ability to prepare in time for problems and opportunities –it is functional in form of preparing for future even unlikely but critical events and not simply predicting what might happen. So context and dynamical representations would be relevant. It could be useful to focus the (cognitive) model rather on the relevant components of e.g. (dynamic) situation understanding and decision making for being able to

trace behavior and cognitive state updating rather than capturing all aspects of cognition (Klaproth et al. 2021, Scharfe et al 2019). For modelling anticipatory ability in a physical HRI further fine-grained social cues must be detected and also produced for realizing joint action. One of the big challenges I see here is to find good ways for information integration and hybrid modelling approaches that are able to meet all the different requirements the individual tasks pose.

Details are given in [1] and [2].

References

- 1 Klaproth, O. W., Halbruegge, M., Krol, L. R., Vernaleken, C., Zander, T. O. and Russwinkel, N. (2020). A Neuroadaptive Cognitive Model for Dealing With Uncertainty in Tracing Pilots' Cognitive State. *Topics in Cognitive Science*, 12(3), p. 1012-1029.
- 2 Scharfe, M. and Russwinkel, N. (2019). Towards a Cognitive Model of the Takeover in Highly Automated Driving for the Improvement of Human Machine Interaction. In *Proceedings of the 17th International Conference on Cognitive Modelling*.

3.20 Anticipatory Brain-Computer Interfaces

Andreea Ioana Sburlea (*University of Groningen, Groningen, NL, a.i.sburlea@rug.nl*)

License © Creative Commons BY 4.0 International license
© Andreea Ioana Sburlea

Intention-related neural signals may be exploited by different human-machine interaction systems. In particular, brain-computer interfaces (BCIs) could be used to restore communication functions and facilitate motor control in paralysed users. BCIs record the brain activity of the user, detect the user's intended action and send the corresponding execution command to an artificial effector system, e.g. computer cursor, prosthetic device, etc. Non-invasive brain recording modalities such as electroencephalography benefit from a fine temporal resolution for the signal acquisition which facilitates the anticipation of different cognitive or motor states in a timely manner.


In my talk, I presented an approach that involves the simultaneous acquisition of multimodal data coming from both neural and behavioral information [1], with the purpose of understanding the underlying mechanisms that link perception to action within the framework of brain-computer interfaces. Specifically, in the context of grasping movements, I showed the importance of disentangling at a neural and behavioral level between the properties of the objects (such as shape and size) and the properties of the movement (such as type of grasp, number of fingers involved) [2]. Moreover, I showed that this disentanglement is similarly observed across modalities in separate stages going from perception, observation to action. Furthermore, I opened the conversation about the influence of the goal and feedback on the perception of action in anticipatory human-machine interaction cases involving different levels of attention to the action.

References

- 1 Sburlea, Andreea I., and Gernot R. Müller-Putz. *Exploring representations of human grasping in neural, muscle and kinematic signals*. *Scientific reports* 8:1 (2018): 1–14
- 2 Sburlea, A. I., Wilding, M., & Müller-Putz, G. R. (2021). *Disentangling human grasping type from the object's intrinsic properties using low-frequency EEG signals*. *Neuroimage: Reports*, 1(2), 100012.

3.21 Anticipation of User Goals from Human Behavioural Data for Human-AI Collaboration

Florian Strohm (University of Stuttgart, DE, florian.strohm@vis.uni-stuttgart.de)

License  Creative Commons BY 4.0 International license
© Florian Strohm

In many machine learning areas, the goal for a model is to solve a specific task faster or better than humans can. However, there are many areas where humans cannot be replaced by an artificial intelligence (AI) system for different reasons, e.g., because humans might not want to be replaced but only seek assistance from an AI, or if the task to solve necessarily includes a human or at least human information. As a consequence, there is an increasing need for AI systems that collaborate with humans, thereby enable them to solve a certain task more efficiently or to achieve a certain goal faster or more easily. For an AI system to collaborate with humans, it is important that the AI understands which information the user is perceiving and processing, as it then can use this information to anticipate user intents. Once the AI system identified user intents it can, depending on the environment and task, anticipate the user's intent, generate results, or perform actions to assist the user. In order to understand which information a human is perceiving and processing it needs to receive information about the user. We specifically focus on using behavioural data as a source of information about the user. The advantage of behavioural data is that the user does not have to explicitly provide it but the AI can automatically obtain user information through their interaction with an application or through external devices like an eye-tracker or using electroencephalography (EEG). [1] One of the main challenges of using implicit behavioural data is the understanding of such data and the subsequent extraction of task relevant information. We plan to develop methods to extract the most relevant information, especially concerning gaze data, e.g., how task relevant a specific fixation is. Since collecting large datasets of such attentive behavioural data is time-consuming and expensive, we will propose methods that allow to generate synthetic datasets and model human behaviour to train intelligent systems.

References

- 1 Strohm, F., Sood, E., Mayer, S., Müller, P., Bâce, M., & Bulling, A. (2021). Neural Photofit: Gaze-based Mental Image Reconstruction. *In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 245-254)*.

3.22 The Role of Skills in Anticipatory Human-Computer Interaction

Niels Taatgen (University of Groningen, NL, n.a.taatgen@rug.nl)

License  Creative Commons BY 4.0 International license
© Niels Taatgen

My specialization is Cognitive Modeling using Cognitive Architectures. I have been involved in ACT-R for a long time, and have contributed mechanisms for production compilation, multitasking and time perception. More recently, I have developed my own variant on ACT-R called PRIMs [1]. The motivation for this was the fact that ACT-R and many cognitive architectures are focused on individual tasks, whereas humans have a history of prior experience that they can use to perform new tasks.

To support models that can perform multiple tasks, I defined an additional level of abstraction between production rules and (task) goals: the level of skills. The idea is that if people have to perform a novel task, they try to mobilize the skills necessary for that task from their available repertoire [2]. Only when they miss particular skills they need to do more extensive learning in order to acquire the missing pieces.

An application area of the skill idea is education, in particular cognitive tutors. Using unsupervised learning methods, we can extract the skill set for a particular domain. Then, give a student's performance on assignments, we can determine which skills the student has mastered, and which not. Based on that, the optimal material or assignment can be selected for the student. In addition, it can be used to provide feedback to both the learner and the teacher.

References

- 1 Taatgen, N.A. (2013). The nature and transfer of cognitive skills. *Psychological Review*, 120(3), 439-471.
- 2 Hoekstra, C., Martens, S., & Taatgen, N. A. (2020). A Skill-Based Approach to Modeling the Attentional Blink. *Topics in Cognitive Science*, 12(3), 1030-1045.

4 Breakout Groups (Monday)

4.1 Shared Representations and Communication

Alvitta Ottley (Washington University in St. Louis, USA, alvitta@wustl.edu)

Matteo Bortoletto (University of Stuttgart, DE, matteo.bortoletto@vis.uni-stuttgart.de)

Nele Russwinkel (TU Berlin, DE, nele.russwinkel@tu-berlin.de)

Oliver Klaproth (Airbus – Hamburg, DE, oliver.klaproth@airbus.com)

Susanne Hindennach (University of Stuttgart, DE, susanne.hindennach@vis.uni-stuttgart.de)

License © Creative Commons BY 4.0 International license

© Alvitta Ottley, Matteo Bortoletto, Nele Russwinkel, Oliver Klaproth, Susanne Hindennach

This breakout group discussed shared representations and communication in anticipatory human-machine interaction.

4.1.1 Shared Representations

Shared representations are often established through discussions/negotiations and can also be reflected in tangible artefacts like sketches/visualisations/mind maps. They comprise representations of task and environment or goals, and can comprise representations of (human/machine) agents and their internal states. The individual representation/understanding of what is shared might differ and such mismatches can be detected and resolved in communication that might require meta-cognition and/or meta-communication.

Conflicts can arise when machines are able to adapt their actions to confirm human anticipation, but know there is a better (optimal) way to achieve the goal. In some situations, the best policy might be to concede, observe and try to learn from human behaviour. In other situations negotiating tasks and goals and providing explanations will be more appropriate. Finding out when to apply what policy might be critical. From an ethical/legal perspective, authority and responsibility should remain with the human where possible.

4.1.2 Communication

Communication is the best way to detect mismatches in representations. We discussed three ways to classify communication.

- **indirect/direct:** Direct communication is about the means of communication and describes interaction without intervening factors to a designated communication partner/audience, indirect communication is interaction with intervening factors.
- **explicit/implicit:** Explicit communication is about the clarity of communication and describes unambiguous interaction, implicit interaction describes ambiguous communication.
- **symmetric/asymmetric:** Symmetric communication describes balance in the use of knowledge, means or tools in communications between two communication partners, asymmetric implies imbalance.

Explicit communication of representations requires effort; implicit detection of mismatches might be more suitable in some situations but challenging. Physiological indicators of representations, intent, etc. can facilitate communication.

4.2 Timescales at which Anticipatory models for HMI can operate

Christian P. Janssen (Utrecht University, NL, c.p.janssen@uu.nl)

Cristina Conati (University of British Columbia – Vancouver, CA, conati@cs.ubc.ca)

Niels A. Taatgen (University of Groningen, NL, n.a.taatgen@rug.nl)

License © Creative Commons BY 4.0 International license
© Christian P. Janssen, Cristina Conati, Niels A. Taatgen

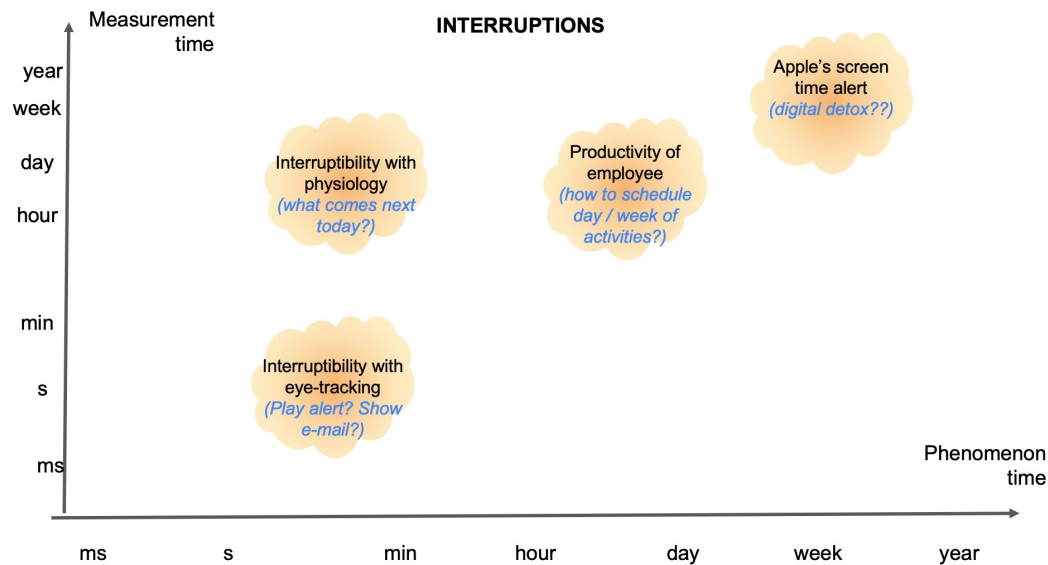
This group thought about the time scales at which anticipatory human-machine interaction could operate. Starting point was Newell’s time scales that distinguish at which time interval behavior (and a model) occurs: milliseconds (biological band), seconds (cognitive band), minutes (rational band), or hours or longer (social band) ([1], see also chapter 1 of [2]).

After some discussion, we thought it might be interesting to identify two dimensions of time: (1) the time dimension at which a phenomenon (such as boredom) occurs, and (2) the time dimension at which a system might measure something about the human to infer user state. These might sometimes be at the same time interval scale, or sometimes differ. In some cases it might be ambiguous what the exact scale is. There might also be other ways to categorise time that our classification does not capture (yet).

Below we have five pictures of four example domains where different scenarios might take up different proportions of these time dimensions. In each figure there are “clouds” that indicate the rough position of the phenomenon and its measurement time. Within that cloud we also added text: the top text describes what the rough phenomenon is, and the bottom text indicates our thoughts about what the phenomenon can be. What is interesting to note about these figures, is that there are opportunities to consider anticipatory systems at many time scales, beyond systems that require immediate action. Although many clouds are positioned close to the diagonal, this should not be interpreted as that these are in principle more common. It might rather show our limitation in thinking of more diverse examples.

References

- 1 Newell, A. (1990). Unified theories of cognition. Cambridge, MA: Harvard University Press.
- 2 Salvucci, D. D., & Taatgen, N. A. (2010). The multitasking mind. Oxford University Press.



■ **Figure 3** Example scenarios related to the domain of interruptions.

4.3 Data-driven vs. rule-based models for anticipation

Jelmer Borst (University of Groningen, NL, j.p.borst@rug.nl)

Andreas Bulling (University of Stuttgart, DE, andreas.bulling@vis.uni-stuttgart.de)

Emmanuelle Dietz (Airbus Central R&T, Hamburg, DE, emmanuelle.dietz@airbus.com)

Chenxu Hao (Friedrich-Alexander-Universität Erlangen-Nürnberg, DE, chenxu.hao@fau.de)

Frank Keller (University of Edinburgh, UK, keller@inf.ed.ac.uk)

Florian Strohm (University of Stuttgart, DE, florian.strohm@vis.uni-stuttgart.de)

Andreea Ioana Sburlea (University of Groningen, NL, a.i.sburlea@rug.nl)

Samuel Kaski (Aalto University, FI & University of Manchester, UK, samuel.kaski@aalto.fi)

Elisabeth André (Augsburg University, DE, elisabeth.andre@uni-a.de)

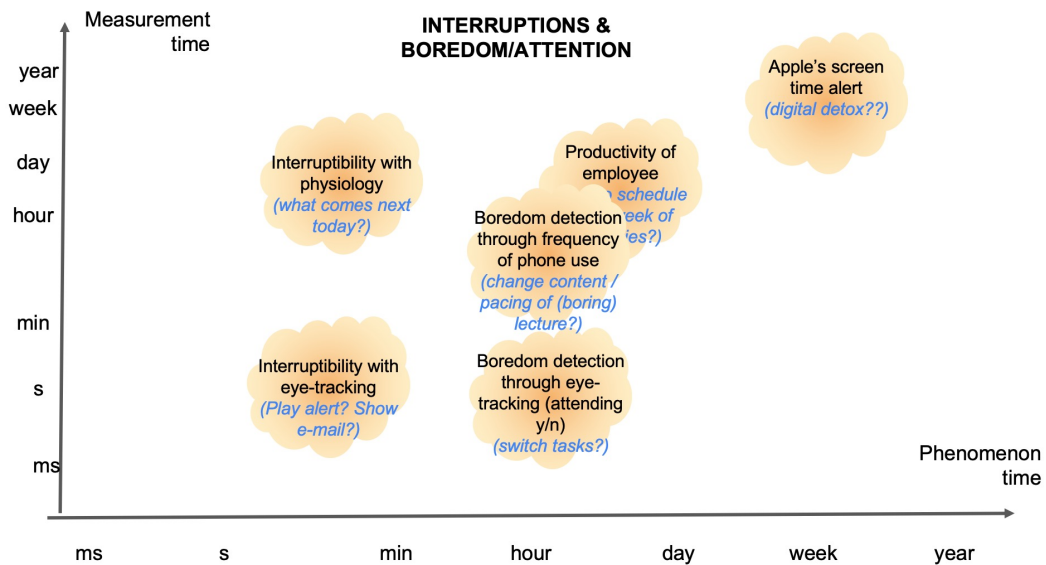
Henny Admoni (Carnegie Mellon University, USA henny@cmu.edu)

License © Creative Commons BY 4.0 International license

© Jelmer Borst, Andreas Bulling, Emmanuelle Dietz, Chenxu Hao, Frank Keller, Florian Strohm, Andreea Ioana Sburlea, Samuel Kaski, Elisabeth André, Henny Admoni

The purpose of this group was to discuss the suitability of rule-based models of anticipation compared to data-driven models. We start with a definition of these terms. Traditional examples of rule-based models include the ACT-R cognitive architecture. However, even ACT-R contains a learning component, e.g., the activation levels are trainable. Other rule-based models includes most models used for planning, and more generally logic-based models. The group felt that some tasks are specified symbolically (e.g., in robotics, where the starting state and the end state of an action is given), and this might bias any modeling effort towards rule-based models. Also, it is important to bear in mind that even rule-driven models are built based on data, except that the model designer comes up with the abstractions (e.g., based on experimental results), and formulates these abstractions in the form of rules.

The prototypical data-driven models are deep-learning models, which are black-box and contain very little internal structure, and are trained on massive datasets. Models with a simple structure are also included, e.g., models that learn state transitions (HMMs,



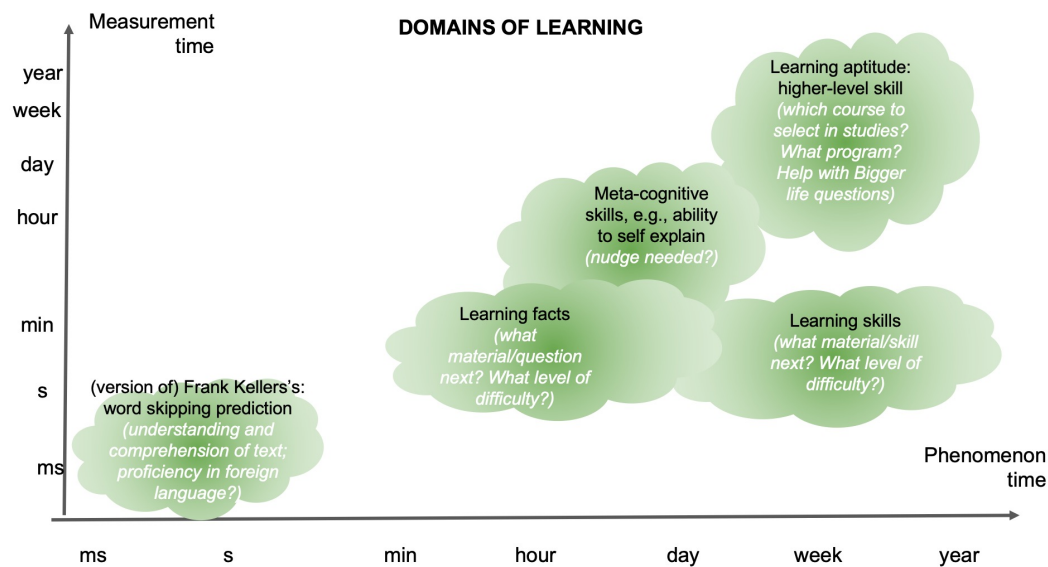
■ **Figure 4** Example scenarios related to the domain of interruptions and boredom.

POMDPs). Instance-based learning is another example. It needs to be taken into account, though, that some data-driven models have internal structure (trees, graphs), for example graph convolutional neural networks.

The group also considered hybrid models and a third model type that combines both data-drive and rule-based models. Hybrid approaches are *prima facie* plausibility, for example because humans probably learn in a hybrid way, combining pre-specified rules (instincts) with data obtained from the environment. Language acquisition is a potential example: the learning mechanism seems to be innate, but external data is required to trigger the acquisition of an individual language. So it seems clear that you can't learn without data, but also not without (possibly pre-defined) structure. A classic example of a hybrid model in machine learning are Bayes Nets, which combine pre-specified (graph) structure with probabilities learned from data.

The group then switched to the discussion of Theory of Mind (TOM) and its relevance to modeling anticipation. As to the definition of TOM, we can think of it as a nested model, i.e., a model within a model. More specifically, having a TOM means having a model of another agent that accounts for their knowledge, beliefs, and intentions. This raises the question of what TOM is useful for and which behaviors we can't do without TOM. Traditionally, non-literal behavior is thought to require TOM: lying, irony, sarcasm, empathy. The group then examined the question whether TOM can be learned in a data-driven way. This is probably possible in specific instances, but generalization would require a rule-based approach. The group also believed that TOM would not be needed for anticipation in the general case; limited, task-specific anticipation behavior could be learned from data. The group also noted that a higher level of abstraction is required to generalize anticipation across tasks, and TOM may be required for that. The link between (impaired) TOM and autism was noted.

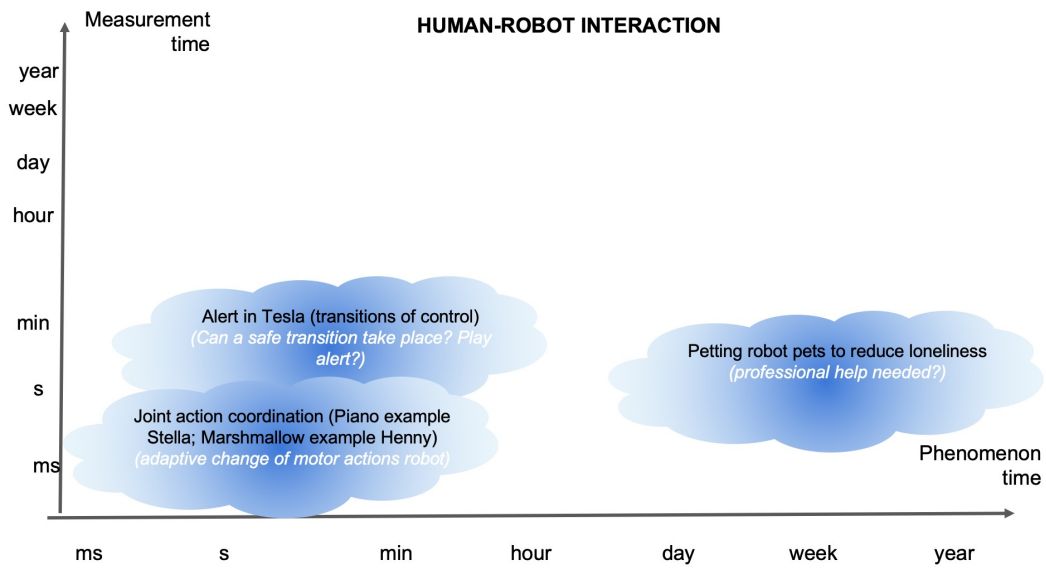
There were other issues that seemed important to the group, but were not discussed in detail because of lack of time. This includes the question of how a simulation is different from a model. The group felt that a simulation required models of both the user and their environment. A simulation could also be used for reasoning, e.g., to simulate what a user



■ **Figure 5** Example scenarios related to the domain of learning.

does in a given environment and therefore reason about what the user would end up doing. Another important topic that the group touched upon was explainability. Humans constantly generate explanations for their behavior and their actions, they tell stories that help them to make sense of the world and to interpret other people and their actions. Another issue the group touched upon was how TOM relates to sense making. Sense making is a way of model criticism, of challenging a model, adapting and modifying it. It was also pointed out that Inference is possible without probabilities, e.g., in Approximate Bayesian Computation (ABC). Another important issue seems to be causality. Can data-driven models be causal? The group felt that this was possible in certain cases, a good example are Bayes Nets.

Finally, the group discussed evaluation. This is potentially an issue with respect to theory of mind: how can you tell that a system has TOM? It may be possible to use tests designed in developmental psychology to test TOM in human participants, e.g., the false belief task (Sally-Anne task). These tests would have to be adapted for a human-machine interaction setting. This raises the wider question of how to evaluate anticipation. How do we tell if a system anticipates successfully? One approach would be to rely on task-based evaluation, i.e., just measure task success. This gives us an indirect way of evaluating anticipation. However, a direct evaluation measure (“anticipation accuracy”) would be desirable, but this remains an issue for future work. Furthermore, the group noted that data-driven models cannot be trained without an evaluation measure that informs the design of the loss function.



■ **Figure 6** Example scenarios related to the domain of human-robot interaction

5 Breakout Groups (Tuesday and Wednesday)

5.1 Formalisation of Anticipation and Use Cases

Cristina Conati (University of British Columbia – Vancouver, CA, conati@cs.ubc.ca)

Emmanuelle Dietz (Airbus Central R&T, Hamburg, DE, emmanuelle.dietz@airbus.com)

Steffen Frey (University of Groningen, NL, s.d.frey@rug.nl)

Susanne Hindennach (University of Stuttgart, DE, susanne.hindennach@vis.uni-stuttgart.de)

Alvitta Ottley (Washington University in St. Louis, USA, alvitta@wustl.edu)

Niels A. Taatgen (University of Groningen, NL, n.a.taatgen@rug.nl)

Matteo Bortoletto (University of Stuttgart, DE, matteo.bortoletto@vis.uni-stuttgart.de)

Henny Admoni (Carnegie Mellon University – Pittsburgh, USA, henny@cmu.edu)

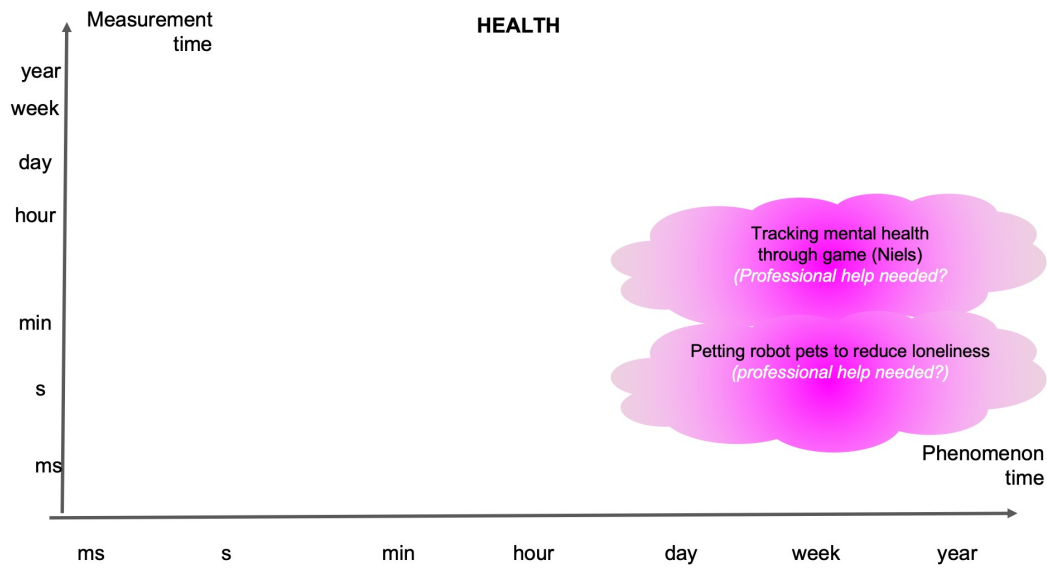
License © Creative Commons BY 4.0 International license

© Cristina Conati, Emmanuelle Dietz, Steffen Frey, Susanne Hindennach, Alvitta Ottley, Niels A. Taatgen, Matteo Bortoletto, Henny Admoni

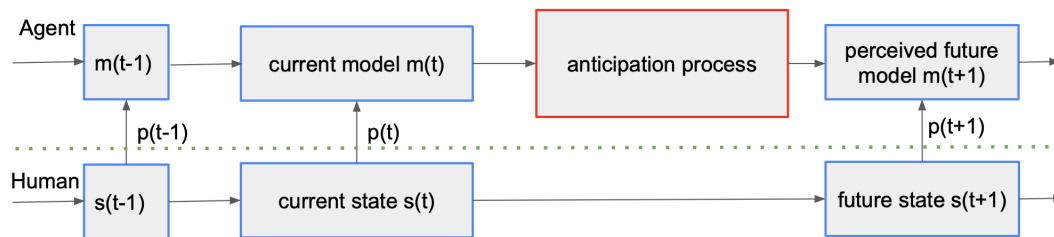
Anticipation is about the future. A system that anticipates can possibly perceive (or recognise) the current (and potentially past) target's state (e.g. human's cognitive state), which determines the prediction of a future state with respect to a set of available actions. Given some objective or utility function (e.g. to prevent a negative outcome or facilitate a positive outcome), the system decides which action to choose (e.g. prepare to act). The system might evaluate the outcome (the perceived future state) in order to evaluate and adapt accordingly.

5.1.1 Notation

The top of Figure 8 (above the green dotted line) represents the perceived environment and the bottom (below the green dotted line) represents the actual (target) environment, i.e. what the system intends to perceive.



■ **Figure 7** Example scenarios related to the domain of health



■ **Figure 8** Overview of the Formalisation of Anticipation.

The models in the perceived environment are from the view of the system (e.g. artificial agent or even a human). The perceived environment is a sequence of models that can change over time $\dots m(t-1), m(t), m(t+1), \dots$ where $t > 1$. The states within the actual environment are actual (target) states (i.e. states that the system intends to perceive), which can range from very simple states (e.g. button on/ off) to very complex states (e.g. the cognitive state of a human). The actual environment is a sequence of states that can change over time, $\dots s(t-1), s(t), s(t+1), \dots$ where $t > 1$. The perceived and the actual environment can differ, depending on the specification of the actual states and the accuracy of the perception function p .

We further introduce the following notation:

- S is a set of states of the world, where each state $s(t) \in S$ is specified with respect to time t ;
- M is a set of models (representations) of the world, where each model $m(t) \in M$ is specified with respect to time t ;
- A is a set of possible actions in the world, where each action $a(t) \in A$ is specified with respect to time t ;
- P is a set of perceptions of the world, where each perception $p(t) \in P$ is specified with respect to time t ;

- $\text{Obj} : M \rightarrow \mathbb{R}$, the objective function, mapping a model m to a value;
- $\rho : S \rightarrow P$, the perception function, mapping a state $s(t)$ to a perception p ;
- $\text{update} : M \times P \rightarrow M$ is the (expectation) update function. It updates a model $m(t)$ given the current perception $p(t)$;
- $\text{prediction} : M \times A \rightarrow M$, the prediction function. It predicts the future model $m'(t+1)$ given the current chosen action $a(t)$ and current model $m(t)$.

The action that give the highest predicted value from model m at time t $\text{mm}(t)$, is computed as follows:

$$a(t) = \underset{a_i \in A}{\text{argmax}} [\text{Obj}(\text{prediction}(m(t), a_i))] \quad (1)$$

where $\text{prediction}(m(t), a_i) \equiv m'(t+1)$ is the predicted effect of action a_i in $m(t)$ and $\text{Obj}(\text{prediction}(m(t), a_i))$ is the predicted value of $m'(t+1)$.

The predicted model according to the selected action producing an *anticipation* is computed by the prediction function *prediction* follows:

$$m'(t+1) = \text{prediction}(m(t), a(t)) \quad (2)$$

The model $m(t+1)$ is computed according to the predicted model $m'(t+1)$ and the perception of state $s(t+1)$ by the update function *update*:

$$m(t+1) = \text{update}(m'(t+1), \rho(s(t+1))) \quad (3)$$

5.1.2 Instantiating the formalism for a particular case

To apply the formalism to a particular problem, the components need to be specified, so we need to provide $\langle M, A, P, \text{Obj}, \rho, \text{update}, \text{prediction} \rangle$. Note that we do not specify S , because the system has no direct access to it.

Each of the individual components can be very complex and the focus of the research, or can be very simple. For example, ρ is simple if all we need to do is check whether a student answered a multiple choice question correctly, or complex when mapping EEG measurements onto workload level. In particular, a lot of variation can be expected in the model M . A model can be not much more than the outcome of perception, but can also be a complete cognitive model with theories about learning, and a record of all past events. Moreover, the model can represent the goal of the system, and a partial plan that is update with experience. An elaborate model will also, in most cases, have more elaborate *update* and *prediction* functions.

5.1.3 Algorithm

Algorithm 1 shows an algorithm applied on a use case about the interaction with a tutoring system. In this task the student interacts with a tutoring system and the systems' goal is that the student gets better at the assignment. Algorithm 2 shows one possible way how the prediction function could be defined for this use case.

Note: The objective function Obj can return how likely an action is to lead to a desired future state or action, that is, how well action a prepares for some goal.

Algorithm 1 Tutoring example.

```

S                                     ▷ what student knows about algebra
A                                     ▷ set of assignments that address specific skills
M = {0, 1}k                           ▷ ability on k algebraic skills
P = {0, 1}                             ▷ correct or incorrect
Obj = ∑ mi
m ← 0k                               ▷ student has no knowledge to start
while Obj(m) < k do
  a ← argmaxa ∈ A(Obj(prediction(m, a)))           ▷ select an action
  m' ← prediction(m, a)                             ▷ make a prediction
  s ← execute a
  observation ← update(m', ρ(s))                       ▷ perceive the new states
  if observation is correct then m ← m'
  else m ← m with some deduction                       ▷ student not as far as we thought
  end if
end while

```

Algorithm 2 Prediction function: predict learning gain given an assignment.

```

Given a ∈ A                               ▷ set of actions (assignments)
Given m ∈ M                               ▷ model (skill levels of the student)
skill ← [0, 1]k                             ▷ which skills are trained by a
δ                                           ▷ learning rate
τ                                           ▷ learning threshold
student_improvement ← ∑ relu{skill(a) − m}     ▷ student skill improvement
if student_improvement < τ then
  m ← m + (δ · skill(a))                       ▷ update student's knowledge
else
  m ← m                                         ▷ expectation is that assignment is too hard
end if

```

5.1.4 Partial/ Uncertain Observations

There will always be some uncertainty as to what is perceived from the system and what is happening in the user (cognitive states). This uncertainty depends on the complexity of what is intended to be perceived. If, for instance, the only requirement is to perceive whether a button is on or off, then the perceived state and the actual state are likely to be the same most of the time (except e.g. if some sensor is broken). However, if the system intends to perceive a human's cognitive state based on the human's behavior, then the perceived state can only provide an approximation.

5.2 Transportation-driven Anticipatory Human-Machine Interaction

Elisabeth André (Augsburg University, DE, elisabeth.andre@uni-a.de)

Mihai Băce (University of Stuttgart, DE, mihai.bace@vis.uni-stuttgart.de)

Jelmer Borst (University of Groningen, NL, j.p.borst@rug.nl)

André Brechmann (Leibniz Institute for Neurobiology Magdeburg, DE, brechmann@lin-magdeburg.de)

Andreas Bulling (University of Stuttgart, DE, andreas.bulling@vis.uni-stuttgart.de)

Christian P. Janssen (Utrecht University, NL, C.P.Janssen@uu.nl)

License © Creative Commons BY 4.0 International license

© Elisabeth André, Mihai Băce, Jelmer Borst, André Brechmann, Andreas Bulling, Christian P. Janssen

5.2.1 A formalism for anticipatory human-machine interaction

We defined anticipatory human-machine interaction as predicting future states, which can include information about human behaviour, the environment, or the system, to prepare for a future machine action.

$$s_{h,t+i} = f(s_{h,t}, a_{m,t+j})$$

where $s_{h,t+i} \in \text{DesiredStates}$, $s_{h,t+i} \notin \text{BlacklistedStates}$, and $t + j < t + i$.

5.2.2 Domain

In our group, the discussions were mainly driven by one key application domain, namely safety-critical human-machine interaction. In particular, our formalism was defined with an aviation use case in mind. Despite a focus on this domain, our formalism can also be applied to a tutoring scenario where a positive intervention could be used to improve the learning process. Finally, we also thought about an application scenario in which the machine schedules interruptions at opportune moments in a task.

5.2.3 Methods

An anticipatory agent has to perceive, reason and react.

Prediction or state tracking is needed to keep track of whether the user has achieved the desired state. Measurement of human / user sensing (perceive) is needed to know the human state, because we want to achieve some desired human state or behaviour.

Learning might be needed to be adaptive to different contexts / settings AND to adjust underlying models in case it (almost) failed in the past (if anticipation was wrong).

We identified a number of methods, according to the different function that they have to fulfil:

- Recurrent networks
- Classifiers
- Anything using Bayesian statistics
- Some form of user / cognitive model

Learning: (can be model-free or model-based) (reason)

- Multi-agent RL
- Symbolic models
- Measurements of human / user sensing (Perceive)
- Physiological + brain

- Question/self-declaration/self-reports
- Observation of actions (taken yes / no? Correct? speed?)

In addition to the above methods that can be used for different abilities of the agent, we discussed possible interruption or intervention modalities.

- Giving a subtle nudge
- Giving a more prominent intervention (such as an alert)
- Brute force taking over behaviour by the system (e.g., transition of control in car)
- Notifying/informing something or someone outside of the system and user (e.g., notifying air traffic control that the pilot seems drowsy, or is in difficult conditions)

5.2.4 Case Study

We discussed the following case study relevant to the aviation domain. Let us assume an unexpected situation in the plane's cockpit, such as a thunderstorm coming up that the pilot may not be aware of. A challenge for anticipatory human-machine interaction: Can the machine (e.g. the plane or a system in the plane) anticipate that the pilot is unaware of the upcoming storm? If yes, how can the system prepare the pilot? The anticipation task in this case it to anticipate the pilot's state of whether they are aware of the storm or not. The machine could suggest alternative routes or request remote support (e.g., air traffic control, etc.).

5.2.5 Challenges

In the breakout session, our group identified a number of key challenges:

- What aspects of user behaviour or state does one need to track, measure, or sense. The state has to be complete (have all the information you need), it should be privacy sensitive (no information you should not need), and feasible to enable reasoning within limited time. Finally, the system should avoid being "over sensitive". Some of the above also hold for systems / machines.
- What are the blacklisted and desired states?
- Act in time before the state is reached, without negatively affecting the user
- Avoid recursive (?) anticipation / be "dynamic" in adjusting actions real-time
- Multi-agent ToM, e.g., system estimates what experienced pilot assumes about state of novice pilot.
- Within aviation domain: there are consequences and systems might not be changed overnight and only once they have been heavily tested. E.g., system needs all kinds of certifications. Transparent, traceable
- Have unambiguous system actions (to avoid that the human/pilot gets distracted by an intervention or misinterpret)


5.3 Planning in Anticipation

Leslie Blaha (AFRL – Wright Patterson, US, lblaha@andrew.cmu.edu)

Matteo Bortoletto (University of Stuttgart, DE, matteo.bortoletto@vis.uni-stuttgart.de)

Samuel Kaski (Aalto University, FI & University of Manchester, UK, samuel.kaski@aalto.fi)

Andreea Ioana Sburlea (University of Groningen, NL, a.i.sburlea@rug.nl)

License  Creative Commons BY 4.0 International license

© Leslie Blaha, Matteo Bortoletto, Samuel Kaski, Andreea Ioana Sburlea

5.3.1 Definition of planning

Planning is the process of thinking regarding the actions to take in order to maximise a form of utility, given a context. Such context can be defined in terms of the following quantities:

- an action space, defining the set of possible actions
- a conditional state space, defining the set of possible states of the environment
- an observation space, defining the set of observations an agent can experience of its environment
- a transition function, giving for each environment state and agent action a probability distribution over environment states
- an observation function, which gives, for each action and resulting state, a probability distribution over possible observations
- a reward function

5.3.2 Methods for planning in Anticipatory Human-Machine Interaction

Two methods are required for planning in Anticipatory Human-Machine Interaction (AHMI). First, we need a method to compute the actual plan. Such method heavily depends on the specific task we are dealing with. Classic approaches involve game theory, probabilistic models, constraint optimisation, causal inference and simulator based inference. Development in deep learning provided other valuable tools, such as active learning and self-play. In particular, combining deep learning methods like self-play and heuristic search algorithms like Monte Carlo tree search led to incredibly powerful models, e.g. AlphaGo. More complex scenarios could require Theory of Mind, e.g. for user modelling. In these cases, inferring humans' mental states could help the AI-assistant to efficiently cooperate with users.

5.3.3 Use cases

The variety of use cases that require (strategic) planning is potentially unlimited. Most decision making scenarios require planning. Some examples are teamwork, model design, education, games or sports, social events, emergency response, crime prevention, rehabilitation, surgery, transportation and logistics.

5.3.4 A formal definition of Anticipatory Human-Machine Interaction in terms of planning

We propose a definition of AHMI based on the concept of planning. *AHMI is helping humans thinking ahead to place resources to maximise eventual utility*, over the space of potential plans. Anticipation is naturally linked to thinking ahead and as we are dealing with human-machine interaction the goal is to help humans. Resources represent some assets that will result on actions taken in order to maximise a form of utility, which is defined by the context.

5.4 Physical Interaction

Nele Russwinkel (TU Berlin, DE, nele.russwinkel@tu-berlin.de)

Frank Keller (University of Edinburgh, UK, keller@inf.ed.ac.uk)

Florian Strohm (University of Stuttgart, DE, florian.strohm@vis.uni-stuttgart.de)

Chenxu Hao (Friedrich-Alexander-Universität Erlangen-Nürnberg, DE, chenxu.hao@fau.de)

Henny Admoni (Carnegie Mellon University, Pittsburgh, USA, henny@cmu.edu)

License © Creative Commons BY 4.0 International license
© Nele Russwinkel, Frank Keller, Florian Strohm, Chenxu Hao, Henny Admoni

Physical human-agent (or human-robot) interaction involves interaction between one or more people and one or more embodied agents in some kind of physical space. This kind of interaction often, but not always, includes the agent having some kind of physical effect in the world. One domain in which physical human-agent interaction has become very salient in recent years is autonomous vehicles (AVs), where drivers, pedestrians, and vehicles must collaborate to navigate safely. For example, a driver may anticipate that the AV's autopilot is likely to fail, and take over control before a critical state is reached. In contrast, a vehicle might anticipate that the driver is available for a control handoff, and yield control of the vehicle. An AV may avoid collisions by anticipating pedestrian actions (e.g., crossing the street), or anticipating the behaviour of other vehicles when merging lanes or performing other negotiations with drivers.

5.4.1 Discussed Problems

Physical human-agent interaction provides a special case of anticipation. Because it involves an embodied agent interacting in the real world, the challenges of the domain can be divided into sensing, planning, and acting. Specifically, some challenges include:

- Sensing
 1. Perception of environment
 2. Perception of user (including social cues)
- Planning
 1. Knowledge of task (e.g., transition function)
 2. Maintaining safety
 3. In collaborative scenarios, also:
 - a. Shared representation
 - b. Bi-directional communication
- Acting
 1. Proactive action
 2. Transparent and explainable action
 3. Adhering to social norms / cues

There are also other challenges to anticipation in physical interactions. These include how to integrate information from multiple input sources, and how to create enough model flexibility to enable generalisation to new situations.

5.4.2 Possible Approaches

Methods for enabling anticipation in physical human-agent interactions centre around enabling bi-directional communication between human and AI agents. This communication can be explicit or implicit, verbal or nonverbal, and direct or indirect.

Direct and explicit communication mechanisms include speech and text. These are generally easy to understand but can be difficult to produce because of their specificity. Implicit behaviours can integrate more seamlessly into ongoing interactions and can add information that supports and augments explicit communication.

For the robot to anticipate the human implicitly, it can sense aspects of human behaviour such as gaze using eye tracking, gestures using motion tracking, and brain signals using brain-computer interfaces or EEG. Other elements that can be sensed from humans include physiological cues and social cues.

Other methods focus on allowing the human to anticipate the robot. To make it easier for humans to do so, robots can generate implicit, interpretable behaviours of their own. These include eye gaze (or head direction), gestures (including pointing), and legible motion. Research has investigated how to automatically generate these behaviours and their effect on human understanding.

To model human and robot in a way that enables physical interaction, robots must be able to represent certain aspects of their environment and partner. These include: a value function to represent reward, a state representation that captures relevant aspects of the world including representations of the human partner(s) and environment, a task representation that allows the robot to plan, and an action space and motion planner that enables the robot to act. Other components of models that are less critical but still useful include an ability to learn or adapt the model, flexibility, and a notion of uncertainty over states and actions.

There are, of course, a number of modeling methods that have been used for anticipation in physical interactions. These include now-standard approaches such as MDP/POMPDs, cognitive models like ACT-R and SOAR, deep neural networks, and Bayes Nets.

5.4.3 Conclusions

Physical human-agent interaction has need of anticipation to ensure safety, task efficiency, and user comfort. Domains like autonomous vehicles or collaborative building require interactions that would benefit from an agent being able to anticipate and respond to human needs, goals, and expectations. The challenges of this domain exist throughout the sense-plan-act cycle. Approaches to address anticipation in physical human-agent interaction must involve representations that enable robot behavior (e.g., state and value functions, motion planners) as well as representations that model the human to be anticipated.

Participants

- Henny Admoni
Carnegie Mellon University –
Pittsburgh, USA
- Elisabeth André
Augsburg University, DE
- Mihai Băce
University of Stuttgart, DE
- Leslie Blaha
Air Force Research Laboratory –
Wright Patterson, USA
- Jelmer Borst
University of Groningen, NL
- Matteo Bortoletto
University of Stuttgart, DE
- André Brechmann
Leibniz Institute for
Neurobiology Magdeburg, DE
- Andreas Bulling
University of Stuttgart, DE
- Cristina Conati
University of British Columbia –
Vancouver, CA
- Emmanuelle Dietz
Airbus Central R&T –
Hamburg, DE
- Steffen Frey
University of Groningen, NL
- Cleotilde Gonzalez
Carnegie Mellon University –
Pittsburgh, USA
- Chenxu Hao
Friedrich-Alexander-Universität
Erlangen-Nürnberg, DE
- Susanne Hindennach
University of Stuttgart, DE
- Christian P. Janssen
Utrecht University, NL
- Samuel Kaski
Aalto University, FI &
University of Manchester, UK
- Frank Keller
University of Edinburgh, UK
- Oliver W. Klaproth
Airbus – Hamburg, DE
- Alvitta Ottley
Washington University in
St. Louis, USA
- Nele Russwinkel
TU Berlin, DE
- Andreea Ioana Sburlea
University of Groningen, NL
- Florian Strohm
University of Stuttgart, DE
- Niels Taatgen
University of Groningen, NL

