Do human anticipation abilities have a special feature?

Anna Strasser – independent researcher

Being able to anticipate future events constitutes a key advantage for any cognitive system. Instead of trial-and-error experiences in the real world, testing hypotheses (in the mind) marks an important difference. If you are able to anticipate the consequences of a potential action, you can avoid risky behavior in the real world. Or as Karl Popper (1984) framed it, one can "*let their hypotheses die in their stead."* This is what Dennett (2018) described as the main difference between so-called Skinnerian and Popperian creatures, modeling different kinds of information-processing that enable competences.

Especially recent research focuses on the competence to anticipate as an essential property by which we can describe brain activity. Namely, the brain is understood as a hypothesis-making system that continually tries to minimize errors in its predictions (see predictive coding / free energy principle: Friston 2009; Hohwy 2013; Clark 2015).

According to Friston, the free energy principle is a universal principle for biological systems. However, assuming such a common anticipatory feature in all biological structures raises the question whether other anticipation systems brought up in our cultural and technological development may be radically different or whether there is a universal core in all anticipatory systems. If one wants to compare human anticipation capacities with those which technological systems have, I would suggest that one can describe at least two dimensions in which human anticipation differs from technological ones.

The first one concerns an essentially *quantitative* dimension. Obviously, the performance of a technological system with AI-enhanced predictive planning differs from the performance of an anticipatory system of an individual human. Artificial systems are able to process and store more data and they do this in a faster way, as compared with humans. That means, with respect to speed and the quantity of processed data, artificial systems easily exceed individual humans. Building upon this quantitative difference, one can observe that human societies make use of technological systems. Therefore, one can assume that human anticipatory systems are influenced and enhanced by the results of technological anticipation systems. For instance, a single person will not be able to anticipate all the details of climate change; but, relying on anticipations brought forward by technological systems, humans can develop more appropriate plans. We can take artificial systems as an extended mind to improve our limited capacities of anticipation (Clark 2015). In sum, such artificial systems with their anticipatory capacities are used as tools.

The second dimension concerns a *qualitative* aspect. Analyzing the type of data which are processed by artificial systems, we might uncover a distinguishing feature of human anticipation. Taking into account that human anticipatory systems fairly seamlessly include social and emotional aspects, one can ask whether artificial systems are as well able to process such data.

With respect to this potential qualitative difference, we should investigate whether the type of data which artificial systems process might entail an important restriction. A lot of successful anticipations in humans rely on socio-cognitive abilities that account for social and emotional data. To anticipate future actions of other agents, it is not only relevant to consider their mental, but also their emotional, states.

The social competence to be able to anticipate what another agent will do serves as a paradigmatic example to investigate the role of social and emotional data processing. In the humanities and natural sciences, one aspect of this social competence is discussed under the label 'mindreading' or 'Theory of Mind' (Fodor 1992). Admittedly, many examples of mindreading studied in the literature tend to exclusively relate to mental states such as knowing and perceiving. Desires are not yet at the foreground of these debates.

Research in artificial intelligence has already demonstrated how artificial agents model mental states of human beings with respect to the perspective of a presumed human counterpart (Gray & Breazeal 2014). Of course, up to now, this is only valid in a limited range of situations. But it shows that artificial agents, in principle, are able to infer from their perception of the physical world whether the human counterpart can see or cannot see an object, and infer that this perspective will guide future actions of the human.

Turning to emotional data, actual research on social robotics is highly relevant, specifically in relation to the development of robots which are designed to enter the space of human social interaction. For example, research about conversational agents has as its objective the development of artificial agents into human-like partners (Mattar & Wachsmuth 2014; Becker & Wachsmuth 2006).

I will not judge whether this direction of research will be successful, i.e., in the sense that artificial systems will eventually be able to bridge the qualitative difference and process all sorts of emotional and social data. However, I will point out certain consequences with which we will be confronted by in case of a success.

In our society using social and emotional data to anticipate behavior of other agents goes along with questions concerning social norms. Consequently, if artificial systems would be successful in processing this type of data, we should approach questions about duties and rights which

artificial agents might then deserve as social interaction partners. On the one hand, it is common sense that artificial agents should not harm humans or other living beings, as, for example, the rules proposed in Asimov's laws about robots (Asimov 1950). On the other hand, it should then also be carefully discussed in what sense artificial social agents may have rights. Since our understanding of fairness and justice is based on how we treat other social agents, it will be essential to develop social norms of how to treat artificial social agents. Furthermore, regarding the outcomes of actions to which artificial agents contribute as social agents, we will have to face new questions about responsibilities. Where previous revolutions have dramatically changed our perspective on the status of women, children, other ethnicities, and animals, this one has the potential to change our understanding of sociality beyond biology. To approach upcoming ethical issues, we need criteria to distinguish social interactions from tool-use. And the ability to process social and emotional data in order to anticipate behavior of other agents might be one of the key features needed for distinguishing tools from genuinely social, but also genuinely artificial, agents.

References

Asimov, I. (1950). *I, Robot.* Gnome Press.

Becker, C. & Wachsmuth, I. (2006). Modeling primary and secondary emotions for a believable communication agent. In D. Reichardt, P. Levi, J. Meyer (eds.), *Proceedings of the 1st Workshop on Emotion and Computing in conjunction with the 29th Annual German Conference on Artificial Intelligence (KI2006), Bremen* (pp. 31–34).

Clark, A. (2015). Embodied Prediction. Open Mind. http://open-mind.net

Dennett, D. (2018). *From Bacteria to Bach and Back: The Evolution of Minds*. W. W. Norton & Company. Kindle-Version.

Fodor, J. (1992). A Theory of the Child's Theory of Mind. *Cognition*, 44(3), 283–296.

Friston, K. (2009). The free-energy principle: a rough guide to the brain?. *Trends in cognitive sciences*, 13(7), 293-301.

Gray, J. & Breazeal, C. (2014). Manipulating Mental States Through Physical Action – A Self-as-Simulator Approach to Choosing Physical Actions Based on Mental State Outcomes. *International Journal of Social Robotics*, 6(3), 315–327.

Hohwy, J. (2013). *The Predictive Mind.* Oxford University Press.

Mattar N. & Wachsmuth, I. (2014). Let's get personal: Assessing the impact of personal information in human-agent conversations. In: M. Kurosu (ed.) *Human-Computer Interaction. Advanced Interaction Modalities and Techniques. Lecture Notes in Computer Science*. Vol 8511. Berlin [u.a.]: Springer, 450-461.

Popper, K. (1984). *Auf der Suche nach einer besseren Welt. Vorträge und Aufsätze aus dreißig Jahren*. München: Piper Verlag.