Hooked on Artificial Agents: A Simulation Study

Ivan Dula[1,4], Tabea Berberena[2], Ksenia Keplinger[3], Maria Wirzberger[1,2,4]

[1] University of Stuttgart, Department of Teaching and Learning with Intelligent Systems, Geschwister-Scholl-Straße 24 D, 70174 Stuttgart, Germany
ivan.dula@ife.uni-stuttgart.de
[2] University of Stuttgart, Stuttgart Research Focus Interchange Forum for Reflecting on Intelligent Systems (IRIS), Geschwister-Scholl-Straße 24 D, 70174 Stuttgart, Germany
tabea.berberena@iris.uni-stuttgart.de
[3] Max Planck Institute for Intelligent Systems, Organizational Leadership and Diversity, Heisenbergstraße 3, 70569 Stuttgart, Germany
kkeplinger@is.mpg.de
[4] University of Stuttgart, Stuttgart Center for Simulation Science, Cluster of Excellence EXC 2075 "Data-Integrated Simulation Science (SimTech)", Pfaffenwaldring 5a, 70569 Stuttgart, Germany
maria.wirzberger@iris.uni-stuttgart.de

With current technological developments and algorithms performing tasks typically reserved for human agents, studies have shown that some people are reluctant when it comes to using and working with artificial agents (Jussupow et al., 2020). This recently identified phenomenon has become known as algorithm aversion (Dietvorst et al., 2018; Castelo et al., 2019; Berger et al., 2021; Chugunova and Sele, 2022). Research, however, is not unified on this issue and some studies have found the exact opposite effect. Algorithm appreciation occurs when humans are appreciative of artificial agents and their advice and may even prefer them to human advice (Jussupow et al., 2020; Chugunova and Sele, 2022; Logg et al., 2019). Furthermore, a phenomenon known as automation bias has been observed where users frequently over-rely on automation, failing to notice errors or discrepancies as well as failing to realize to intervene when they should (Chugunova and Sele, 2022). Many factors influencing which of the three types of behavior will emerge in a given situation have been identified. They include transparency, control, social influence, trust, and others. What is still missing from the literature is a unifying foundational model capable of explaining these behaviors on a structural level. This is crucial if we want to have a good understanding of the impact that the implementation of artificial agents will have on our organizations as well as leverage points that can help steer our systems in a desired direction.

One way to address this need is by using systems thinking. Systems thinking can be described as a perspective, a language, and a set of tools, through which it is possible to make sense of the complexity of the world by looking at it in terms of wholes and relationships, rather than splitting it down into its parts (Kim, 1999; Ramage and Shipp, 2009). It has been applied to a wide range of fields and disciplines due to its ability to solve complex problems, explain non-linear behaviors, understand socio-economic problems, and understand seemingly illogical behaviors of individuals, countries, and organizations (Monat and Gannon, 2015). It focuses on eliciting the system structure as a

way of increasing our understanding of the observed behavior. Only when we understand the structure and the relationship between structure and behavior can we begin to understand how systems work, what makes them produce poor results, and how to shift them into better behavior patterns (Meadows, 2008).

The first step in eliciting system structure and attempting to understand system behavior is drawing causal loop diagrams (CLD's). CLD's are one of the most important tools of systems thinking which enable us to capture how variables in a system are interrelated. They take the form of one or more closed loops that depict cause-and-effect linkages. These loops indicate the presence of reinforcing and/or balancing processes, which determine the behavior of dynamic systems. Once completed, we will deploy another valuable systems thinking tool, called systems archetypes. In systems thinking, archetypes are common problem-causing structures that are repeated in many situations, environments, and organizations (Monat and Gannon, 2015). Currently, there are 10 common archetypes and identifying them in dynamic systems is the first step towards changing destructive structures and behaviors (for a detailed overview of systems archetypes see Kim and Anderson, 1998). Our goal is to identify the common structural causes behind automation bias, algorithm aversion, and algorithm appreciation, in order to produce a unifying theoretical foundation for these significantly different observations, as well as identify the underlying systems archetype(s) in order to better understand the sources of problematic behaviors and the important leverage points to mitigate them. Finally, we will develop a quantitative simulation model capable of reproducing problematic behaviors in order to conduct a series of simulation experiments with the goal of identifying the main drivers of the successful implementation of artificial agents in organizations.

References

[1] Berger, B., Adam, M. Rühr, A., and Benlian, A. 2021. Watch Me Improve - Algorithm Aversion and Demonstrating the Ability to Learn. *Business and Information Systems Engineering, 63*(1), 55–68. https://doi.org/10.1007/s12599-020-00678-5

[2] Castelo, N., Bos, M.W., and Lehmann, D.R. 2019. Task-Dependant Algorithm Aversion *Journal of Marketing Research, 56*(5), 809-825. https://doi.org/10.1177/0022243719851788

[3] Chugunova, M. and Sele, D. 2022. We and It: An interdisciplinary review of the experimental evidence on how humans interact with machines. *Journal of Behavioral and Experimental Economics, 99*(2022), 101897. https://doi.org/10.1016/j.socec.2022.101897

[4] Dietvorst, B.J., Simmons, J.P., and Massey, C. 2018. Overcoming Algorithm Aversion: People Will Use Imperfect Algorithms If They Can (Even Slightly) Modify Them. *Management Science, 64*(3), 1155–1170. https://doi.org/10.1287/mnsc.2016.2643

[5] Jussupow, E., Benbasat, I., and Heinzl, A.. 2020. Why are we averse towards algorithms? A comprehensive literature review on algorithm aversion. *Research papers*,168. https://aisel.aisnet.org/ecis2020-rp/168/

[6] Kim, D.H. 1999. Introduction to Systems Thinking. Pegasus Communications, Inc., Waltham, Massachusetts. www.pegasuscom.com

[7] Kim, D.H. and Anderson, V. 1998. *Systems Archetype Basics: From Story to Structure.* Pegasus Communications, Inc., Waltham, Massachusetts. www.pegasuscom.com

[8] Logg, J.M., Minson, J.A., and Moore, D.A. 2019. Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes, 151*(2019), 90-103. https://doi.org/10.1016/j.obhdp.2018.12.005

[9] Monat, J.P. and Gannon, T.F. 2015. What is Systems Thinking? A Review of Selected Literature Plus Recommendations. *American Journal of Systems Science, 4*(1), 11-26. https://doi.org/10.5923/j.ajss.20150401.02

[10] Ramage, M. and Shipp, K. 2009. *Systems Thinkers*. Springer London, United Kingdom. https://doi.org/10.1007/978-1-4471-7