Adaptive Design in Human-Robot Collaboration*

Paul M. Gorny, Louis Schäfer

Abstract

Artificial intelligence (AI) and sensor technology have made collaborative robots (cobots) more relevant in the industry. However, an understanding of how human-cobot interaction differs from human-human interaction is vital, as cobots become more adaptive to workers' behaviour and needs. We conducted a controlled and incentivized "field-in-the-lab" experiment in a realistic production environment. In all treatments, there was a human Worker 1 at Station 1 in a two-station production line. We varied whether Worker 2 at Station 2 was a human or a robot. The teams produced electronic motor components, and Worker 1 could submit intermediate products to Worker 2 or a "waiting queue." All components that completed the queue or were finished by Worker 2 counted toward the team payoff. Worker 1's production speed was classified as "high" (H), "medium" (M), or "low" (L). In line with this speed, in the "adaptive" treatments, the robot was set to production speed H>M>L, Worker 2 received productivity feedback H>M>L, or Worker 2 received piece rate H>M>L. We find strong robotic aversion and blame shifting toward the robots. Adaptivity increased productivity and mildly reduced robotic aversion. It also reduced the allocation of blame, but this reduction was not statistically significant. This study has important implications for using cobots and can inform the adaptive design of effective human-robot collaboration, which can enhance workplace productivity.

Keywords: robots, blame, algorithm aversion

JEL Classifications: C91

*Gorny: Institute of Management (IBU), Karlsruhe Institute of Technology. Schäfer: Institute of Production Science (wbk), Karlsruhe Institute of Technology. We would like to thank Nimet Bozkurt, Niklas Busse, Sergiu Panainte, Haythem Rebhi, Tom Reuscher, Karoline Ströhlein, Niklas Tremmel, Sven Walther, Fabian Wüst, Rebecca Zimmer, and Annika Zsohar for excellent research assistance. The authors gratefully acknowledge funding from the IFREE Small Grants Program.

Collaborative robots (cobots) become more common in industrial settings (Graetz and Michaels, 2018; Cheng et al., 2019). Thus, it is crucial to understand how human-cobot interaction differs from human-human interaction. Cobots are designed to work safely alongside and in close physical proximity to human workers, allowing the cobots and human workers to interact. Cobots can thus be a valuable workforce addition, especially for repetitive, dangerous, and physically more strenuous tasks. On the one side, this means human-cobot interaction will become more frequent in industrial settings. Thus, it is important that human workers accept the cobots as coworkers and that they are willing to actively include them in the production process. On the other side, the focus on safe interaction implies that the social impact of the human-cobot interaction in the workplace is not the primary focus in cobot design. Thus, it is also important to understand whether the cobots' inclusion in the production process induces a diffusion of responsibility with the cobots receiving more blame for negative work outcomes than the human coworkers. Human workers might feel an inhibition to blaming their human coworkers–a behavioural convention potentially not (yet) applying when interacting with a cobot. We thus want to investigate whether (i) human workers prefer human coworkers over cobots when working on a joint task and (ii) whether human workers have a lower inhibition to shifting blame to a cobot as compared to shifting it to a human coworker, when the blame allocation has real (monetary) consequences.

To answer the first question, we start out from previous findings that suggest that a "lack of control" and a perceived "black box nature" of algorithms, artificial intelligence (AI), and machines more generally can lead to lower acceptance of algorithms, or even to aversion (Dietvorst et al., 2015, 2018; Dietvorst and Bharti, 2019). However, AI has become increasingly adaptive in recent years to react to worker behaviour. As a result, cobots are now increasingly being operated using AI technology, mainly to enhance their safety when reacting to what workers do, but consequently also increasing their capabilities for adaptivity. In a broad definition, adaptivity in terms of worker behaviour refers to the ability of a system or machine to adjust its behavior based on the actions and preferences of the worker. If workers realise their influence on the cobots through this channel, this could give them an increased feeling of control. Therefore, it is crucial to investigate whether the adaptive design of human-cobot interactions can increase robot acceptance or reduce robot aversion. For our second research question, we investigate whether this change in perceived control also leads to a change in perceived responsibility and, in turn, a change in blame allocation.

We conducted a "field-in-the-lab" experiment (Kandler et al., 2021) to investigate human-robot collaboration in a controlled, incentivised, and realistic production environment. In all treatments, we set up a production line with two stations and there always was a human participant in the role of Worker 1 at Station 1. We varied whether Worker 2 at Station 2 was a human participant or a robot.[1] These production teams produced an electronic motor component. Worker 1 could submit intermediate products to Worker 2 or a neutral "waiting queue." All components that completed the waiting queue–after a certain time elapsed–or were finished by Worker 2 counted toward the payoff of all human workers in the team. We used a team piece rate with 60 ECU for each counted component at an exchange rate of 1 ECU=€0.01. Worker 1's production speed was classified to be "**H**igh," "**M**edium," or "**L**ow." In the *Human-Static* and *Robot-Static* treatments, this had no effect on Worker 2. In the Adaptive treatments, in line with Worker 1's production speed, (a) the robot was set to a production speed H>M>L (*Robot-Adaptive*), (b) Worker 2 received productivity feedback H>M>L (*Human-Feedback*), or (c) Worker 2 received piece rate H>M>L (*Human-Incentive*). Worker 1 was informed about the setup in each treatment. We tested whether this adaptivity affected Worker 1's behaviour. Subsequently, we elicited blame in an incentivised decision. We analysed the ratio of intermediate products passed to team members relative to the waiting queue to test for "robotic aversion" and allocation of blame between team members.

We find strong evidence for both robotic aversion and a higher blame allocation to the robotic compared to the human coworkers. Adaptivity had a strong positive effect on productivity. Controlling for team productivity, we find mild evidence for more intermediate products being passed to the robotic coworker in the *Robot-Adaptive* than the *Robot-Static* treatment. We also find mild evidence that participants passed more intermediate products to human coworkers in the adaptive (*Human-Feedback* and *Human-Incentives*) than the *Human-Static* treatment. Thus, adaptivity increases the degree to which Worker 1 integrates Worker 2 into the production process (irrespective of whether Worker 2 is a human or robotic coworker). Regarding the allocation of blame, adaptivity qualitatively reduces the amount of blame shifted to Worker 2–again irrespective of whether Worker 2 is a human or robotic coworker–but this difference is not statistically significant.

We mainly relate to two strains of the literature. One is the economics and psychology literature of human-machine interaction (HMI), mainly using lab and field experiments. Chugunova and Sele (2020) and Jussupow et al. (2020) provide excellent literature surveys on the more recent studies within this methodological framework. Recent studies that have received particular attention are those relating to the phenomena of *algorithm aversion* (Dietvorst et al., 2015, 2018; Dietvorst and Bharti, 2019) and *algorithm appreciation* (Logg et al., 2019). The aforementioned literature surveys suggest that aversion is more pronounced in moral and social domains, whereas appreciation (and lower aversion) is more likely to be found when people have some degree of control over the

[1] We used a KUKA KR 6 R900 sixx.

automated agent. Savela et al. (2021) report evidence from a vignette study suggesting that humans in mixed human-robot teams have a lower in-group identification than those in purely human teams. Leo and Huh (2020) suggests that participants would attribute more blame to themselves than to a robot. This connects the literature on HMI to the economic literature on blame shifting (Oexl and Grossman, 2013; Feier et al., 2022), that we also inform with our experiment.

Our study has important implications for the design and implementation of cobots in the workplace. Adaptive design of effective human-robot collaboration needs to consider that humans may prefer to work with human team members over cobots. Adaptivity can play an important role in moderating such preferences. Additionally, it is important to recognise that humans may allocate blame differently between human and robot team members. With an increasing amount of 'agentic' technology (AI, chatbots, robots/cobots) being employed in the work context, this could become a barrier to personal and social responsibility.

References

Cheng, H., R. Jia, D. Li, and H. Li (2019): "The rise of robots in China," *Journal of Economic Perspectives*, 33, 71–88.

Chugunova, M. and D. Sele (2020): "We and it: An interdisciplinary review of the experimental evidence on human-machine interaction," *Max Planck Institute for Innovation & Competition Research Paper*, 20.

Dietvorst, B. J. and S. Bharti (2019): "Risk Seeking Preferences Lead Consumers to Reject Algorithms in Uncertain Domains," *ACR North American Advances*.

Dietvorst, B. J., J. P. Simmons, and C. Massey (2015): "Algorithm aversion: people erroneously avoid algorithms after seeing them err." *Journal of Experimental Psychology: General*, 144, 114.

——— (2018): "Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them," *Management Science*, 64, 1155–1170.

Feier, T., J. Gogoll, and M. Uhl (2022): "Hiding behind machines: Artificial agents may help to evade punishment," *Science and Engineering Ethics*, 28, 19.

Graetz, G. and G. Michaels (2018): "Robots at Work," *The Review of Economics and Statistics*, 100, 753–768.

Jussupow, E., I. Benbasat, and A. Heinzl (2020): "Why are we averse towards algorithms? A comprehensive literature review on algorithm aversion," European Conference on Information Systems.

Kandler, M., L. Schäfer, P. M. Gorny, G. Lanza, P. Nieken, and K. Ströhlein (2021): "Learning factory labs as field-in-the-lab environments–an experimental concept for human-centred production research," in *Proceedings of the Conference on Learning Factories (CLF)*.

Leo, X. and Y. E. Huh (2020): "Who gets the blame for service failures? Attribution of responsibility toward robot versus human service providers and service firms," *Computers in Human Behavior*, 113, 106520.

Logg, J. M., J. A. Minson, and D. A. Moore (2019): "Algorithm appreciation: People prefer algorithmic to human judgment," *Organizational Behavior and Human Decision Processes*, 151, 90–103.

Oexl, R. and Z. J. Grossman (2013): "Shifting the blame to a powerless intermediary," *Experimental Economics*, 16, 306–312.

Savela, N., M. Kaakinen, N. Ellonen, and A. Oksanen (2021): "Sharing a work team with robots: The negative effect of robot co-workers on in-group identification with the work team," *Computers in Human Behavior*, 115, 106585.