Proceedings of the 35th European Safety and Reliability & the 33rd Society for Risk Analysis Europe Conference
Edited by Eirik Bjorheim Abrahamsen, Terje Aven, Frederic Bouder, Roger Flage, Marja Ylönen

doi: 10.3850/978-981-94-3281-3_ESREL-SRA-E2025-P3927-cd

Human-Robot Collaboration for Industry 4.0: Managing Risk and Enhancing Performance Through Mutual Adaptation

Popolo Valentina

Università Telematica Pegaso – Naples , Italy Email: valentina.popolo@unipegaso.it

De Rosa Francesco

Department of Chemical, Materials and Industrial Production Engineering, University of Naples Federico II, Naples, 80125, Italy E-mail: francesco.derosa22@studenti.unina.it

Granitto Raffaele

Department of Chemical, Materials and Industrial Production Engineering, University of Naples Federico II, Naples, 80125, Italy E-mail: ra.granitto@studenti.unina.it

Rizzo Vincenzo

Department of Chemical, Materials and Industrial Production Engineering, University of Naples Federico II, Naples, 80125, Italy E-mail: vin.rizzo@studenti.unina.it

Abstract: Technological progress in the industrial context has encouraged using robots and autonomous systems within industries, focused on introducing solutions to monitor the safety and risks related to Human-Robot Collaboration (HRC). The proposed analysis will meet this requirement, using a mathematical model, which, starting from the systematic analysis of traditional interaction models, analyses the mutual interplay between robots and humans. It will evaluate the impact of this collaboration, quantify mutual learning, consider human feedback to improve the adaptation of cobots to the working environment, and manage human unpredictability, which is responsible for new risk factors. The goal is to improve performance, reduce operator stress and the risk of accidents in the workplace and consequently increase safety.

Keywords: Industry 4.0, Human-Robot Collaboration, Risk Analysis, Mutual Learning, Safety.

1. Introduction

The growing interaction between humans and robots within industrial environments makes it necessary to address the problem of managing safety and the risks associated with them. This integration brings new problems, since robots operate in environments shared with humans, and the dynamics between these two actors are not always predictable.

The evolution of industrial production depends on adopting robots that can perform repetitive actions and interact with humans to accomplish more complex functions to increase productivity and flexibility (Zhao et al., 2024). Human-Robot Collaboration (HRC) is a central matter of interest in the innovative factory sector (Liu & Wang, 2018).

Recent studies have highlighted the benefits of HRC, raising the strengths of robots and humans to achieve higher production rates; in particular, robots are expected to perform more challenging and monotonous tasks while humans deal with more complex and innovative ones (Arents et al., 2021). Companies aim to modify work environments to create harmonious coexistence between humans and robots, looking for efficiency and safety in the workplace (Nicora, Ambrosetti, & Wiens, 2020).

Reviewing the HRC models, their shortcomings will be highlighted, and the key areas for improvement and implementation will be outlined.

The main problem is the need to guarantee that this collaboration safely takes place, minimizing the risks of accidents and improving the operators' working conditions. The main challenge comes from complexity being fundamentally connected to robots' aptitude for following predefined rules and patterns; humans can behave unpredictably, creating new risk situations. This leads to difficulty adapting cobots to the work environment, especially when responding in real time to human behaviour (Wang, Wan, Li, & Zhang, 2016).

Therefore, the objective is to find a solution that allows monitoring this interaction and constantly improving safety in the workplace. This model will not only consider a static model but will also take human feedback into account, trying to improve robots' adaptability to the working environment and better manage human unpredictability.

The main problem in HRC concerns the management of memory and learning over time, which partly determines the safety of operators. An individual's memory decreases over time without any improvement interventions, as highlighted by the Ebbinghaus forgetting curve. The proposed model aims to consider the rate of memory decay, which reflects how quickly information is forgotten, and the mutual adaptation between humans and cobots, which can increase the operator's real and perceived safety to increase trust in the robot.

The dominant challenge is to integrate a dynamic synergy between human and robot through a series of feedback, per the law of practice of Newell and Rosenbloom. The primary difficulty lies in adapting the behaviour of robots in real time, which follow preset patterns. For legal reasons, now cannot yet be equipped with artificial intelligence in response to human unpredictability, improving safety in the workplace and optimizing efficiency (Zirar et al., 2023).

2. Literature Review

The proposed model represents a significant advancement in HRC, integrating adaptive variables such as dynamic trust, task complexity, and digital twin support. This combination enables more flexible and proactive HRC, meeting the needs of Industry 4.0 and paving the way for Industry 5.0, which aims at even more natural and sophisticated human-robot cooperation (Li et al., 2023).

One of the main topics of Industry 4.0, which attempts to incorporate automation, robots, and artificial intelligence into industrial processes, is the development of HRC. However, robotics integration with humans in shared work environments presents significant safety, risk management, and mutual learning performance improvement issues (Rahman et al., 2023).

Theoretical models that describe human learning processes and cognitive decline provide the basis for a sizable portion of the literature on Human-Machine Interaction (HMI). The Ebbinghaus forgetting curve (1885) and the Newell and Rosenbloom law of practice (1980) are the most notable. Although both models provide important insights, they are severely limited regarding dynamic interaction with intelligent technologies.

The Ebbinghaus curve, as shown in Figure 1, illustrates how human learning tends to deteriorate dramatically over time in the absence of reinforcement. Although this paradigm is crucial for understanding memory processes, it disregards external factors such as machine interface and active feedback. For example, a robot might act as a support agent in an industrial setting by providing constant and customized reinforcement, slowing down the rate at which human memory deteriorates.

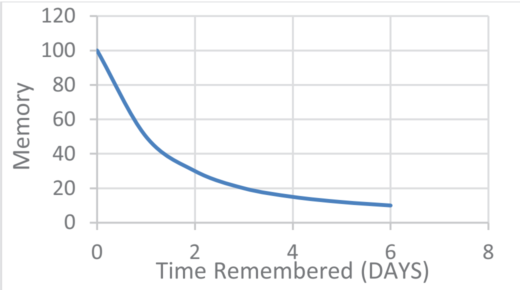

Figure 1- Ebbinghaus Curve

According to a law proposed by Newell and Rosenbloom, human performance rises rapidly in the early stages of learning before levelling off. This hypothesis emphasizes the idea that progress slows down with experience. The static model does not consider the dynamic synergy between man and robot, which is necessary to enhance performance in a collaborative setting.

One of the main goals of contemporary HRC research is to decrease the likelihood of accidents. Recent studies have shown that humans can handle complex situations by having robots perform dangerous and repetitive tasks (Bolla et al., 2023). However, these remedies are incompatible with human behaviour's unpredictable nature, a significant risk factor in communal settings.

Traditional HRC models focus on using gesture recognition to enable communication between humans and robots (Baratta et al., 2024). Reactive robots can react to human inputs in real time, but their capacity for ongoing adaptation is frequently constrained (Borghi et al., 2025). However, these approaches do not include reciprocal learning between people and robots. They specifically do not dynamically describe how human feedback enhances robot replies or how synergy slows down human cognitive decline.

The proposed model will incorporate a human-robot synergy element, unlike the Ebbinghaus curve, which presumes a passive and unstoppable memory deterioration. This feature allows to use the robot's continuous input to postpone the decline of human intellect to improve learning in the early stages of interaction when the human is more receptive to its suggestions and corrections.

The following research questions will be discussed in Methodology paragraph:

1. What specific mathematical framework will be utilized to evaluate the risks inherent in HRC?
2. How will the proposed model accommodate the inherent variability in human behaviour during real-time interactions with robotic systems?
3. What strategies will be adopted to facilitate mutual learning and optimize memory retention between humans and cobots over an extended period?

3. Methodology

Starting from a few established models about HMI, such as those described in literature reviews, Hermann Ebbinghaus's "Forgetting Curve" and Allen Newell & Paul Rosenbloom's "Law of Practice" have created an integrated model that aims to overcome each one's weakness and fill the gap in the literature.

Opening with a description of these models, the discussion will proceed to their integration, highlighting the innovative aspects of each and the limitations overcome in the dynamic analysis.

Ebbinghaus' theory, as anticipated, represents, through his forgetting curve, one of the fundamental contributions in the history of psychology and the study of memory, describing the decline of memory over time, showing how learned information is quickly forgotten immediately after learning, with a gradual decrease in the rate of forgetting over time (Ebbinghaus, 1885).

Ebbinghaus's revolutionary idea was to study memory under controlled conditions, eliminating factors such as meaning and pre-existing associations that could influence the results. To this end, he developed an innovative method based on nonsense syllables and consonant-vowel-consonant combinations to minimize the connection with common language (Roediger, 2015). With this and other strategies, such as the saving method, he created a quantitative method on the forgetting process, which has been taken up and analysed by many scholars over the years who have described it mathematically through a relationship. Ebbinghaus's law of forgetting, used to represent the natural decay of human learning over time (Murre & Dros, 2015), is expressed as follows:

$$M(t) = M_0 * e^{-\gamma t} \qquad (1)$$

In the suggested framework, this law is represented by the components:

$$L(t) = A_0 * e^{-\beta t} \qquad (2)$$

Table 1 explains the main terms used by Ebbinghaus's methodology:

	Nomenclature Table
$M(t)$	Retention level or memory time at time t: an individual's ability to retain information or skills over time after learning them.
M_0	Initial level of memory or learning at time t=0
γ	Decay rate of memory: how quickly knowledge is lost over time without reinforcing interventions.
t	Time elapsed since learning or the starting point of observation.
$L(t)$	Learning level at time t: which includes memory and human-robot synergy.
β	Decay rate of learning adapted to the context of HMI
A_0	Initial learning level at time t=0

Table 1- Ebbinghaus Nomenclature

This term represents the human component of natural decay, modelled exactly like the Ebbinghaus curve.

Allen Newell and Paul Rosenbloom's theory is based on the concept that improvement in performance on a complex task, as experience increases, follows a power law relationship. When a person begins practising a new task, the improvements are rapid. This means the learning curve is quite steep at first. After gaining a certain amount of experience, improvements become less noticeable. The curve flattens, and further improvement becomes more difficult even if the practice continues (Newell & Rosenbloom, 1980).

A common mathematical formulation of the law of practice is:

$$P(t) = P_0 * t^{-n} \qquad (3)$$

To adapt it to our model, we implemented the human-robot synergy component as follows:

$$1 + \alpha * e^{-\delta * t} \qquad (4)$$

This term adds a dynamic feedback effect to the model. At the start of the learning or interaction process, when t is small, the exponential term is close to 1, and the synergy effect is high. As t increases, the exponential term decays, and the influence of the machine's feedback diminishes.

In essence, this term shapes how much cobot contributes to improving the human's performance over time. Its impact is most pronounced initially and taper off as the human becomes more proficient and less dependent on the machine's guidance.

Table 2 explains the main terms used by Newell and Rosenbloom's methodology:

	Nomenclature Table
$P(t)$	Performance at time t represents the level of skill or accuracy achieved after a certain amount of practice or repetition.
P_0	Initial performance at the beginning of practice (t=0) is usually the baseline skill level before practising or learning.
n	A positive exponent determines how quickly the improvement slows down. The larger β is, the faster the initial improvement occurs, but the

	improvement slows down more quickly over time.
t	Time or Trial Number: The number of practice sessions, trials, or time elapsed during the learning or practice process.
α	A constant that determines the strength of the synergy effect or feedback. The larger the value of α, the greater the cobot positive influence on the learning or retention process.
δ	A constant that controls how quickly the feedback's influence diminishes over time.

Table 2- Newell & Rosenbloom Nomenclature

The initial synergy $1 + \alpha * e^{-\delta * t}$ (4) is maximal at the beginning but decreases over time in an exponentially decreasing fashion, modelling the concept of slowing improvement with practice, consistent with the Power Law (performance improves more slowly over time). This component represents the effect of robot support and mutual feedback that slows decay and accelerates human learning in the early stages.

Combining the two formulas, the law that explains the objective of the formulated method has been showing, which aims to describe how human learning varies when collaborating with a cobot as a function of time. It integrates the following criteria:

1. Natural decay, without external support, human learning decays exponentially over time, as described by Ebbinghaus, contained in the term $A_0 * e^{-\beta t}$

2. In the human-robot synergy effect, robot provides positive feedback, temporarily improving the human's performance and slowing the decay. This effect is maximum at the beginning and decreases over time following an exponential decreasing trend, modelled by the Power Law of Practice:$1 + \alpha * e^{-\delta * t}$.

3. Dynamic combination, human memory and robot contribution multiply, creating a synergistic interaction that dynamically varies over time.

$$A(t) = A_0 * e^{-\beta t} * \left(1 + \alpha * e^{-\delta * t}\right) \qquad (5)$$

It is didactically relevant to highlight the differences between R(t) and A(t), underline the improvements made in our model, and highlight the impact of feedback and collaboration with the cobot, which can slow down the natural decay of memory.

These differences are summarized in Table 3:

	M(t) (Retention Level)	A(t) (Learning Level)
Definition	Indicates the amount of memory or knowledge retained over time.	Represents the overall learning level, which includes memory and human-robot synergy.
Origin	Ebbinghaus' forgetting curve.	Human-robot synergy model, integrating feedback and decay.
Formula	$M_0 * e^{-\gamma t}$	$A(t) = A_0 * e^{-\beta t} * \left(1 + \alpha * e^{-\delta * t}\right)$

Context	Focuses on the natural memory retention of an individual, which decays over time.	Considers learning in a dynamic context involving collaboration with a cobot.
Sinergy Component	Does not include interaction with a robot.	Includes the positive effect of robot feedback.
Decay	Decays exponentially over time in a natural manner.	Decays exponentially but is slowed down by human-robot synergy.

Table 3 – Differences between R(t) and A(t)

In practice, M(t) provides a pure representation of memory decaying over time without the influence of external factors, while A(t) gives a complete and more dynamic picture of human learning when supported by technological tools.

The starting models—Ebbinghaus, Newell, and Rosenbloom—exhibit application limits addressed in the developed framework.

The Ebbinghaus forgetting curve is a passive learning curve. It does not consider the influence of feedback or active collaboration, which could improve the memory rate or retentivity over time. Therefore, it represents rapid and static forgetting without considering external interactions.

The new model overcomes this by introducing dynamic feedback and a collaborative factor between humans and cobots. It overcomes this limit by allowing humans to learn more quickly and reducing forgetting since robots can adapt and improve responses based on the user's learning. It can also become a partner that helps maintain learning, improving long-term memorization and not only providing static information.

The Newell and Rosenbloom model lacks mutual adaptation. It does not consider the dynamic adaptation between humans and robots. For example, if robot becomes better at responding in a personalized way based on humans' needs, the improvement may not be linear.

The proposed model, by introducing the factor $1 + \alpha * e^{-\delta * t}$ (4), dem onstrates how col laboration bet ween humans and cobots can increase the speed of learning even in the advanced stages of the process.

As a result, robots and humans do not learn separately, but by adapting to each other, they continue to improve together, overcoming the stagnation predicted by the power law.

3.1 Case of use

To clarify the concepts, it will be presented a numerical application that will show the learning laws found as t varies; it will be replaced "t" in the main formula (5) as follows:

$$A(t) = A_0 * e^{-\beta t} * \left(1 + \alpha * e^{-\delta * t}\right)$$

- Initial phase (t=0):

For t=0, human memory ($A_0 * e^{-\beta t}$) is at its maximum ($= A_0$), and the robot contribution $\left(1 + \alpha * e^{-\delta * t}\right)$ is at its maximum (1+1=2). Learning is greatly improved by collaboration with it, mathematically, too.

- Intermediate stage (medium t):

For an intermediate t, the human memory decay term ($e^{-\beta t}$) begins to dominate. The robot continues to contribute to slowing the decay, but its impact diminishes with time.

- Advanced stage (high t):

Human memory is almost completely decayed for high t tending to infinity, and the robot contribution becomes negligible ($e^{-\delta * t} \sim 0$).

Overall, learning is reduced to minimal levels, practically zero.

3.2 Proposition

Subsequent studies and future research will propose implementing this model with a human stress and fatigue monitoring-based addition. This will enable the simulation of the impact of stress on individual performance and allow for a comparison with synergistic learning under cobot assistance. By pseudocode, implementing this model in Matlab will enable visualizing how cobots reduce cognitive overload, promoting industrial productivity and operator well-being.
The framework that governs these aspects is expected to have a form like this:

$$S_{total} = \alpha * S_{stress} + \beta * D_{dynamics} + \gamma * E_{emotional} + \delta * R_{robot} \quad (6)$$

The variables are defined as follows:

- S_{total} is the total stress of the operator.
- S_{stress} is stress related to the operator's physiological and psychological state, such as heart rate or blood pressure variation during the interaction with the robot.
- $D_{dynamics}$ represents the group dynamics, including cooperation, communication, and conflict among team members (both humans and robots).
- $E_{motional}$ is the operator's emotional response, such as anxiety, frustration, or satisfaction, derived from collaboration with the robot.
- R_{robot} reflects the robot's behavior and its impact on the operator, such as predictability, reliability, and autonomy.

The parameters $\alpha, \beta. \gamma, and\ \delta$ in the theoretical function are scalar weights that determine the relative importance of each factor in the overall function S_{total}. Weights can be calculated by analyzing data collected from empirical studies or simulations. For instance, by measuring how stress changes as each factor varies or by normalization, weights are generally normalized $(\alpha + \beta + \gamma + \delta = 1)$ to ensure that the total sum reflects a balanced distribution.

4. Conclusion

In conclusion, HRC in industrial environments represents an excellent opportunity to improve production efficiency. However, it also entails significant risks related to the unpredictability of human behaviour and the decay of human memory. The proposed model, which combines the Ebbinghaus forgetting curve with dynamic robot feedback and the law of practice, innovatively addresses these issues. It allows memory decay to slow down through a synergy between man and robot, improving performance and safety. The introduction of dynamic mutual feedback empowers a smoother and safer collaboration, responding adaptively to the operator's needs and increasing the system's reliability. This approach closes the gaps of traditional models and paves the way for evolution towards an Industry 5.0 model, where the partnership between man and cobot becomes even more natural and efficient, with positive impacts on productivity and operator well-being.

References

Arents, J., Abolins, V., Judvaitis, J., Vismanis, O., Oraby, A., & Ozols, K. (2021). Human-robot collaboration trends and safety aspects: A systematic review. *Journal of Sensor and Actuator Networks,* 10(3),48. https://doi.org/10.3390/jsan10030048

Baratta, A., Cimino, A., Longo, F., & Nicoletti, L. (2024). Digital twin for human-robot collaboration enhancement in manufacturing systems: Literature review and direction for future developments. *Computers and Industrial Engineering,* 187,109764. https://doi.org/10.1016/j.cie.2023.109764

Bolla, R., Bruschi, R., Davoli, F., Lombardo, C., Mohammadpour, A., Trivisonno, R., & Poe, W. Y. (2023). Adaptive reliability for the automated control of human-robot collaboration in beyond-5G networks. *IEEE Transactions on Network and Service Management,* 20(3),2489–2503. https://doi.org/10.1109/TNSM.2023.3298171

Borghi, S., Ruo, A., Sabattini, L., Peruzzini, M., & Villani, V. (2025). Assessing operator stress in collaborative robotics: A multimodal approach. *Applied Ergonomics,* 123,104418. https://doi.org/10.1016/j.apergo.2024.104418

Ebbinghaus, H. (1964). *Memory: A contribution to experimental psychology* (H. A. Ruger & C. E. Bussenius, Trans.). New York: Dover. (Original work published 1885).

Li, S., Zheng, P., Liu, S., Wang, Z., Wang, X. V., Zheng, L., & Wang, L. (2023). Proactive human–robot collaboration: Mutual-cognitive, predictable, and self-organising perspectives. *Robotics and Computer-Integrated Manufacturing,* 81, 102510. https://doi.org/10.1016/j.rcim.2022.102510

Liu, H., & Wang, L. (2018). Gesture recognition for human-robot collaboration: A review. *International Journal of Industrial Ergonomics,* 68, 355–367. https://doi.org/10.1016/j.ergon.2018.07.005

Murre, J. M. J., & Dros, J. (2015). Replication and analysis of Ebbinghaus' forgetting curve. *PLoS ONE,* 10(7), e0120644. https://doi.org/10.1371/journal.pone.0120644

Newell, A., & Rosenbloom, P. S. (1980). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1–55). Hillsdale, NJ: Erlbaum.

Nicora, M. L., Ambrosetti, R., & Wiens, G. J. (2020). Human-robot collaboration in intelligent manufacturing: Reactive behavioral intelligence of robots. *L'Industrie,* 143, 031009.

Rahman, Md. S., Ghosh, T., Aurna, N. F., Kaiser, M. S., Anannya, M., & Hosen, A. S. M. S. (2023). Machine learning and internet of things in Industry 4.0: A review. *Measurement: Sensors,* 28, 100822. https://doi.org/10.1016/j.measen.2023.100822

Roediger, H. L., III. (2015). Remembering Ebbinghaus. *Contemporary Psychology: A Journal of Reviews,* 30(7), 519–523. https://doi.org/10.1037/023895

Wang, S., Wan, J., Li, D., & Zhang, C. (2016). Implementing smart factories of Industry 4.0: An outlook. *International Journal of Distributed Sensor Networks,* 12(1),1–10. https://doi.org/10.1177/1550147715629101

Zirar, A., Ali, S. I., & Islam, N. (2023). Worker and workplace Artificial Intelligence (AI) coexistence: Emerging themes and research agenda. *Technovation, XX*,102747. https://doi.org/10.1016/j.technovation.2023.102747

Zhao, Y., Said, R., Ismail, N. W., & Hamzah, H. Z. (2024). Impact of industrial robot on labour productivity: Empirical study based on industry panel data. *Innovation and Green Development,* 3, 100148. https://doi.org/10.1016/j.igd.2024.100148