26th International Congress on Modelling and Simulation, Adelaide, SA, Australia, 30 Nov–4 Dec 2025
mssanz.org.au/modsim2025

Minimising Ergonomic Impact in Automated Workspaces through Reinforcement Learning

R. García-Flores [a], Gabriel Dennis [b], Daniel Smith [c], Sevvandi Kandanaarachchi [a], Ashfaqur Rahman [c] and Simon Harrison [a]

[a] *CSIRO Data61, Private Bag 33, Clayton South, VIC, Australia, 3168*
[b] *CSIRO Data61, 1 Technology Court, Pullenvale, QLD, Australia, 4069*
[c] *CSIRO Data61, 15 College Road, Sandy Bay, TAS, Australia, 7005*
Email: Rodolfo.Garcia-Flores@csiro.au

Abstract: Collaborative robots (or *cobots*) are emerging as a transformative technology in industry because they add agility and flexibility to manufacturing processes. Unlike traditional robotics, the field of cobotics focuses on the cooperative interaction between humans and machines. However, the ergonomic impact on the human operators of working in spaces shared with automated or semi-automated mechanical aids is not fully understood, even though increased digitalisation of workspaces provides an opportunity to quantify productivity and risks. In the present paper, we propose an evidence-based methodology to produce an optimal strategy (i.e., a policy) to minimise the ergonomic impact over human operators while they produce a part of a vehicle in close proximity to machines. We model the partially controlled worker's ergonomic states as a Markov Decision Process (MDP) and optimise the policy using Reinforcement Learning (RL) through value iteration. Actual worker's data is collected using *Ergomechanic*, a software tool that conducts non invasive collection of ergonomic data from individuals using Markerless Motion Capture (MMC) technology. Provided all workplace and ethical constraints are respected, the continuous observation of the interaction of a human worker with mechanical aids provides solid evidence to optimise workers' safety on the factory floor. Our results show that the effect of modifying incentives to continue (or not) working under fatigue produce different safety policies. While the values that these incentives should take to avoid injury are not obvious, they can be justified by using the evidence collected by MMC, opening the door to quantitative health and safety studies of human-machine interactions.

Keywords: *Reinforcement Learning*, *Ergonomics*, *Markov Decision Process*, *Markerless Motion Capture*, *Policy Optimisation*

1. INTRODUCTION

Unlike traditional industrial robots, which carry out their assigned tasks independently and without human intervention, collaborative robots, or *cobots*, assume the presence of a human operator to perform their tasks. Cobots are purposefully designed for direct interaction with people and perform their tasks in close proximity to human operators. Cobots, which may take the form of exoskeletons or remotely controlled manipulators, are guided by users and are meant to help the operator accomplish the task at hand by complementing movements, by increasing the force exerted, or by compensating the weight of an object or a tool [Bounouar et al., 2022].

The ergonomic impact of the new work regime brought about by cobots is still poorly understood. Safety in hybrid environments (i.e., shared by humans and machines) relies on sensors to limit their speed and force, and on the use of soft and lightweight materials for their construction. However, beyond these general protections, there exist few specific design guidelines for Health and Safety (HS) in these environments, and none of them is based on continuous monitoring and measurement of human-machine interactions. On one hand, it is known [Safe Work Australia, 2021] that overuse musculoskeletal injuries are the most prevalent type of workplace injury, with large associated social and economic cost but, unfortunately, it is often impractical to collect body load data [Harrison et al., 2025], making it difficult for engineers and researchers to gather evidence and optimise workspace design. On the other hand, there are standards such as ISO 15066 [ISO, 2011] and others that provide guidelines on safety monitoring stops, hand-guiding, speed and separation monitoring, and force-limiting as preventative measures. However, the use of cobots is not widespread enough as to provide the evidence to develop more specific, ad-hoc policies to ensure the safety and wellbeing of the human workforce.

In the present paper, we address these gaps and present an evidence-based methodology for the optimal design and operation of hybrid workspaces. First, we propose the collection of data from a Markerless Motion Capture (MMC) system. The system we use is CSIRO's *Ergomechanic*[1], which is a software tool that conducts non-invasive collection of ergonomic data from individuals using MMC technology without the need of wearable sensors. Ergomechanic extracts the joint positions of a participant from a video feed, from which several ergonomic metrics can be computed to determine if the participant is at risk of developing an overuse musculoskeletal injury. Then, with the data obtained, we produce a Markov Decision Process (MDP), which is a stochastic model describing a sequence of possible events. We assume that the transitions of the the MDP are partially controlled by the worker's *actions*, so that we can optimise a decision policy to minimise the risk of injury. The proposed MDP mimics the situation on safe industrial workspaces. The methodology assumes that workspace configuration affects the probability of a worker being fatigued, injured, or suffering an incident while on duty.

1.1. Previous Work

Optimisation studies for cobotic robot arms on their own are relatively common, e.g., How et al. [2023], optimised the control scheme of a robot hand that can obtain real-time human motion information. Of more interest are the applications of Markov Chains MC to hybrid workspaces and HS. Regarding the use of MC in ergonomics, Jeong et al. [2019] used MC to explore whether driving safety in situations involving curved roads and secondary tasks can be evaluated using multiple measures of eye movement. Specifically, the MC was used to quantify the drivers' dynamic eye movement patterns, in addition to typical static visual measures, providing insights into the subject's strategies. Schmid [2022] employed Bayesian Exploratory Factor Analysis (BEFA), first introduced by Conti et al. [2014], to analyse the data of a flight simulation study regarding single pilot operations. The author used a series of Markov Chain Monte Carlo models, needed by BEFA, to assess the latent factor structure of a framework known as the Technology Acceptance Scale for the two different crewing conditions and their corresponding workstation and cockpit setups for the copilot. Veditti et al. [2018] extended the Faulty Tree Analysis methodology with fuzzy sets and MC, and demonstrated its application with an example of using a press brake.

[1] https://www.csiro.au/en/work-with-us/ip-commercialisation/marketplace/ergomechanic, accessed on February 06, 2025.

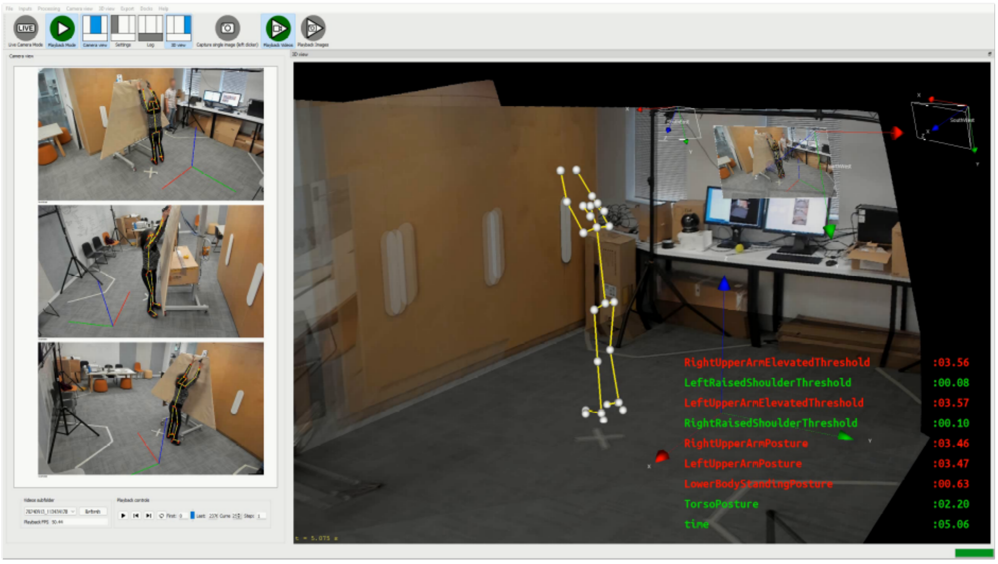

Figure 1: The Ergomechanic human motion capture tool. The subject is working on a part mounted in a trolley. The 3D model of the body is shown as a yellow stick figure, and ergonomic metrics are shown in the bottom right. Part and trolley can be seen clearly on the left panel.

Regarding the assessment of probabilities of injury or fatigue, there exist a number of methodologies, although the vast majority of them seem to be based on surveys and coded in Health and Safety manuals. Di Bona et al. [2021] combined three existing methodologies to propose a new one called *Logit Human Reliability*. Additionally, this reference lists performance-shaping factors and the dependencies between them. Ideally, we would like these probabilities to be derived from the automated collection of workers' motion data at the factory floor as captured by Ergomechanic [e.g., in Cohen et al., 2020]. This topic is of interest to the present paper so that the transition probabilities of the MDP can be determined reliably.

Sobhani et al. [2015] developed a cost optimisation model that incorporates a two-state MC to quantify injury risk factor effects and estimate their economic impacts in optimising a serial assembly line's performance. The MC considers the employees' health-related productivity loss and fits the model using data from government sources and insurance agencies. Even though Sobhani et al. [2015] stressed the importance of accounting for the indirect effects of human factors, these authors still presented a model whose objective function is expressed in terms of costs. Sobhani and Wahab [2017] extended the approach by using a similar model with a three-state MC to model the effects of ergonomic conditions on the carbon emissions of a manufacturing system. It is worth noting that Sobhani et al. [2015] and Sobhani and Wahab [2017] used MCs to model changes in human health states, whereas our intention in this project is to use the probabilities of change between these health states to ultimately *improve the workers' safety.*

In summary, Markov analysis and optimisation can contribute many insights into a worker's behaviours and motion patterns, including cognitive processes at work and hazardous behaviours. This brief review of existing work demonstrates that, despite its potential, the use of MC in ergonomics and workspace design is still rare.

2. METHODOLOGY

2.1. Data Collection Through MMC

Figure 1 shows a still image from a video of a volunteer working on a part mounted on a trolley, as captured by Ergomechanic. A reasonable use of Ergomechanic on a factory floor, provided all workplace and ethical constraints are respected, is to automatically produce a log of incidents (Table 1) over a period of time, listing the time series information of the worker's ergonomic

Table 1: An example of an automatically produced Health and Safety log with information on the ergonomic state of the worker stored every 15 minutes.

start_times	end_times	ergonomic_state	incident	action
2020-01-01 09:00:00	2020-01-01 09:15:00	usual	FALSE	none
2020-01-01 09:15:00	2020-01-01 09:30:00	usual	FALSE	rest (break)
2020-01-01 09:30:00	2020-01-01 09:45:00	usual	FALSE	none
...	...	...	...	...
2020-01-01 16:30:00	2020-01-01 16:45:00	usual	FALSE	none
2020-01-01 16:45:00	2020-01-01 17:00:00	usual	FALSE	rest (end_of_day)
2020-01-02 09:00:00	2020-01-02 09:15:00	usual	FALSE	none
...	...	...	...	...
2020-01-03 11:30:00	2020-01-03 11:45:00	usual	FALSE	none
2020-01-03 11:45:00	2020-01-03 12:00:00	fatigued	TRUE	none
2020-01-03 12:00:00	2020-01-03 12:15:00	fatigued	TRUE	none
...	...	...	...	...

state. The log may be expected to contain, at a minimum, the beginning and end of a time interval, the ergonomic state that the worker is in, and if a HS incident such as a near miss was recorded. It may also capture other actions, such as when the worker takes a short break or goes on leave. These are shown under column *action*. It is assumed that the log contains enough information to produce a *probability transition matrix* of the worker's states.

It is worth noting that, at present, *Ergomechanic* can extract the joint positions and angles of a participant from a video feed and can also compute *ergonomic metrics* from the ISO 11226 standard to determine if there is a risk of developing overuse musculoskeletal injury. Describing the conversion mechanism from these metrics into a categorical state of wellbeing as listed in column `ergonomic_state` of Table 1 is outside the scope of this paper, but could be achieved by different methods, such as by setting thresholds on the metrics, or by using Functional Data Analysis (FDA). FDA is commonly used to analyse processes that are better described as curves of continuously monitored data or smooth functions. The worker's movements in time, described as a time series, can be used as data points, and fatigue and injury can then be identified as outliers.

2.2. Estimating an Optimal Ergonomic Policy

The HS log contains information about the likely changes from one ergonomic state to another (i.e., transitions), and the influence that other actions could have in those changes. Because actions have a partial influence on the transitions between ergonomic states, this could be framed as a decision problem where an optimal strategy (or policy) that minimises risks and ergonomic impact can be estimated. The transitions between ergonomic states can be modelled using a discrete Markov Chain, and the decision problem as a MDP, where transitions are partially controlled by selecting *actions*. These are introduced next and solved by RL in Section 4.

3. MATHEMATICAL MODEL

3.1. Markov Decision Process

Figure 2 shows the *Markov decision Process* for the ergonomic impact minimisation problem. The variables that describe the dynamics of this decision process are:

1. The set of states $\mathcal{S} = \{u, n, f, i\}$, denoting usual, near miss, fatigue and injury, respectively.

2. The set of actions $\mathcal{A} = \{r, w\}$, denoting rest and work.

3. The transition matrix $T(s_{t+1}|s_t, a)$ that defines the probability of reaching state s_{t+1} given the current state s_t and action a. If action a can be taken while in state s_i, $\sum_j T_{ij} = 1$, otherwise the sum is zero. Rest returns the worker to the usual state with certainty, whereas

the work transition probabilities are split.

$$T(s_{t+1}|s_t, r) = \begin{array}{c c} & \begin{array}{cccc} u & n & f & i \end{array} \\ \begin{array}{c} u \\ n \\ f \\ i \end{array} & \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \end{array}; \; T(s_{t+1}|s_t, w) = \begin{array}{c c} & \begin{array}{cccc} u & n & f & i \end{array} \\ \begin{array}{c} u \\ n \\ f \\ i \end{array} & \begin{pmatrix} 0.9921 & 0.0014 & 0.0064 & 10^{-04} \\ 0.5385 & 0.0000 & 0.4615 & 0.0000 \\ 0.0000 & 0.0000 & 0.9818 & 0.0182 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 \end{pmatrix} \end{array}.$$

4. The reward function $R(s_t, a, s_{t+1})$ that returns a scalar reward based on reaching next state s_{t+1} from state s_t after taking action a. The assumed non-zero rewards for this process are $R(u, w, u) = 10$, $R(f, r, u) = 5$, $R(n, w, u) = 1$, $R(n, r, u) = 1$ (which are positive and therefore encouraged), $R(u, w, i) = -10$, and $R(f, w, f) = -2$ (which are discouraged).

This MDP is obtained from the log in Table 1 and the method presented in Barratt and Boyd [2023] to fit feature-dependent Markov Chains.

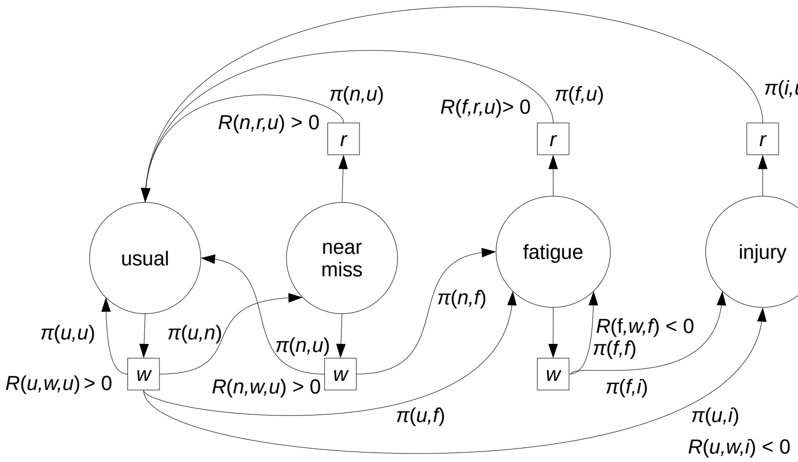

Figure 2: The Markov Decision Process. A transition between states (circles) is represented by an arrow. The two available actions (denoted by squares) are *rest* (r) and *work* (w). An action influences, but does not guarantee, a transition.

3.2. Reinforcement Learning

RL is a framework for teaching agents to learn and make decisions in their environments. Algorithm 1 describes RL's solution approach by value iteration. First, the value function V is initialised to zero for all states in $\mathcal{S}$. Then, the change in value estimates across all states Δ is set to zero (Step 4). We set variable v to the value of state s, $V(s)$ (Step 6), calculate the value of the Bellman equation to update $V(s)$ by weighting all available actions moving forward from each state (Step 7) and update Δ (Step 8) until the convergence criterion is met (Step 10). A more detailed description of the algorithm may be found in [Brandimarte, 2021].

4. RESULTS

The following results were obtained in a Dell OptiPlex 7090 desktop with a 12-core Intel i5-10500 at 4.5 GHz running R version 4.5.1 and library `ReinforcementLearning` version 1.0.5 [Proellochs and Feuerriegel, 2020] under Ubuntu 22.04.5. The execution time is approximately 15 seconds with a simulated log of $100,000$ entries. Table 2 shows the optimal policy π for the base case calculated with discount factor $\gamma = 0.8$ and threshold $\theta = 0.001$, and for varying values of $R(f, w, f)$, which is the reward to continue working when fatigued. Although the only valid actions in states u and i are w and r, respectively, it is not obvious what the worker's action should be in case of near miss or fatigue, and hence the use of RL provides insights on the incentives, using as evidence the worker's actual behaviour on the factory floor as captured by Ergomechanic's MMC. For example, Table 2 shows that increasing the incentive $R(f, w, f)$ from

Algorithm 1 – Reinforcement Learning through value iteration

Input: Sets of states $\mathcal{S}$ and actions $\mathcal{A}$; transition T and reward R matrices.
Initialise $V(s)$ to zero, for all $s \in \mathcal{S}$.
repeat
  $\Delta \leftarrow 0$.
  for `each` $s \in \mathcal{S}$ **do**
    $v \leftarrow V(s)$.
    $V(s) \leftarrow \max_a \sum_{s_{t+1},r} p(s_{t+1}, r|s_t, a)\, [r + \gamma\ V(s_{t+1})]$.
    $\Delta \leftarrow \max(\Delta, |v - V(s_t)|)$
  end for
until $\Delta < \theta$.
Output: A policy $\pi \approx \pi^*$, such that $\pi(s) \leftarrow \arg\max_a \sum_{s_{t+1},r} p(s_{t+1}, r|s_t, a)\, [r + \gamma\ V(s_{t+1})]$.

Table 2: Optimal policies for the MDP in Figure 2. The shaded row is the base case.

$R(f, w, f)$	**State:**	injured	fatigued	near miss	usual
-2	**Policy:**	rest	rest	rest	work
9	**Policy:**	rest	work	rest	work
11	**Policy:**	rest	work	work	work

-2 (a safe base case) to 9 produces the policy $\pi = (r, w, r, w)$, and increasing it further to 11 results in (r, w, w, w). These changes would occur if the worker was under increased pressure to continue working, even when fatigued, at the expense of increased risk of injury. The log table and MDP used here to demonstrate the minimisation of ergonomic impact are simple, but they could be expanded to include states of other mechanical aids present in the hybrid workspace, and their ergonomic effects over workers.

5. CONCLUSIONS AND FURTHER WORK

Cobots are emerging as a transformative technology in industry, but the health and safety implications of their use on human operators is still poorly understood. Even when there are guidelines and standards regarding their use, few studies have been published that rely on evidence gathered on the factory floor. In the present paper, we have proposed a methodology that uses actual worker's movement data collected via MMC as evidence to produce an optimal strategy (or policy) that minimises the ergonomic impact over human operators while they work manually in spaces that are shared with mechanical aids. The worker's data is used to determine the probabilities of transition of a MDP, and then used to calculate the optimal policy using RL's value iteration algorithm. Although the long-term effect of fatigue may be cumulative (i.e. non-Markovian), the model is a reasonable first approximation assuming normal work shifts.

Even when the structure of the log and the MDP are relatively simple, the results show that our approach is feasible and that variations on the incentives, which depend on a worker's individual traits and organisational culture, can be modelled effectively. Most importantly, we show that the probabilities and incentives can be quantified by a MMC system like Ergomechanic, opening the door to the quantitative measurement of the impact of adding cobots to the work environment. Future work includes the addition of cobot's states and the development of models of human-machine interactions.

REFERENCES

Barratt, S. and S. Boyd (2023). Fitting feature-dependent Markov chains. *Journal of Global Optimisation 87*(2–4), 329–346.

Bounouar, M., R. Bearee, A. Siadat, and T. Benchekroun (2022, October). On the role of human operators in the design process of cobotic systems. *Cognition, Technology and Work 24*(1), 57–73.

Brandimarte, P. (2021). *From Shortest Paths to Reinforcement Learning.* EURO Advanced Tutorials on Operations Research. Springer.

Cohen, R., S. Harrison, and P. Cleary (2020, September). Dive Mechanic: Bringing 3D virtual experimentation using biomechanical modelling to elite level diving with the workspace workflow engine. *Mathematics and Computers in Simulation 175*, 202–217.

Conti, G., S. Frühwirth-Schnatter, J. Heckman, and R. Piatek (2014). Bayesian exploratory factor analysis. *Journal of Econometrics 183*(1), 31–57.

Di Bona, G., D. Falcone, A. Forcina, F. De Carlo, and L. Silvestri (2021, November). A hybrid model to evaluate human error probability (HEP) in a pharmaceutical plant. In *Proceedings of the 17th IFAC Symposium on Information Control Problems in Manufacturing (INCOM)*, Volume 54, pp. 444–449. IFAC Papers Online.

Harrison, S., R. Cohen, J. O'Hanlon, T. Cece, N. Fisher, and P. Thrall (2025). Proof-of-concept system evaluation of Ergomechanic for non-invasive estimation of upper-body posture and body load exposure in the workplace. *International Journal of Occupational Safety and Ergonomics*, 1–13. https://doi.org/10.1080/10803548.2025.2458442.

How, J., Y. Jiang, L. Zhao, P. Yu, and J. Pian (2023, May). Research on master-slave teleoperation control algorithm based on exoskeleton master hand. In *Proceedings of the 2023 35-th Chinese Control and Decision Conference, CCDC*, Yichang, People's Republic of China, pp. 828–833. IEEE.

ISO (2011, July). Robots and robotic devices — Collaborative robots. https://www.iso.org/standard/62996.html.

Jeong, H., Z. Kang, and Y. Liu (2019, August). Driver glance behaviors and scanning patterns: Applying static and dynamic glance measures to the analysis of curve driving with secondary tasks. *Human Factors and Ergonomics in Manufacturing and Service Industries 29*(6), 437–446.

Proellochs, N. and S. Feuerriegel (2020). *ReinforcementLearning: Model-Free Reinforcement Learning.* R package version 1.0.5, https://github.com/nproellochs/ReinforcementLearning.

Safe Work Australia (2021). Work-related traumatic injury fatalities, Australia 2021. Technical Report ISSN 2209-9190, Safe Work Australia. www.safeworkaustralia.gov.au/, accessed on March 5 2024.

Schmid, D. (2022, January). The technology acceptance scale: Its Bayesian psychometrics assessed in a factor analysis via Markov Chain Monte Carlo models. *Human Factors and Ergonomics in Manufacturing and Service Industries 32*(1), 66–88.

Sobhani, A. and M. Wahab (2017, November). The effect of working environment-ill health aspects on the carbon emission level of a manufacturing system. *Computers and Industrial Engineering 113*, 75–90.

Sobhani, A., M. Wahab, and W. Newman (2015, March). Investigating work-related ill health effects in optimizing the performance of manufacturing systems. *European Journal of Operations Research 241*(3), 708–718.

Veditti, T., N. Tran, and A. Ngo (2018, July). Dynamic fuzzy safety analysis of an industrial system. In P. Arezes (Ed.), *Advances in Safety Management and Human Factors, AHFE 2017*, Volume 604 of *Advances in Intelligent Systems and Computing*, Los Angeles, pp. 465–471. Springer.

