Journal of Intelligent & Robotic Systems (2022) 106:45
https://doi.org/10.1007/s10846-022-01747-5

SHORT PAPER

High Performance on Atari Games Using Perceptual Control Architecture Without Training

Tauseef Gulrez[1] · Warren Mansell[2]

Received: 14 September 2022 / Accepted: 17 September 2022 / Published online: 8 October 2022

Abstract

Deep reinforcement learning (DRL) requires large samples and a long training time to operate optimally. Yet humans rarely require long periods of training to perform well on novel tasks, such as computer games, once they are provided with an accurate program of instructions. We used perceptual control theory (PCT) to construct a simple closed-loop model which requires no training samples and training time within a video game study using the Arcade Learning Environment (ALE). The model was programmed to parse inputs from the environment into hierarchically organised perceptual signals, and it computed a dynamic error signal by subtracting the incoming signal for each perceptual variable from a reference signal to drive output signals to reduce this error. We tested the same model across three different Atari games Breakout, Pong and Video Pinball to achieve performance at least as high as DRL paradigms, and close to good human performance. Our study shows that perceptual control models, based on simple assumptions, can perform well without learning. We conclude by specifying a parsimonious role of learning that may be more similar to psychological functioning.

Keywords Perceptual control theory · Atari games · Machine learning · Artificial agent

1 Introduction

Gaming environments are increasingly used to develop artificial intelligence. In order to play in a gaming environment, deep reinforcement learning (DRL) agents require a long training time to learn and execute commands. The most successful model-free DRL agents DQN [18], A3C [19], and Rainbow [9] which achieved human-level performance on Atari games, go through a trial-and-error training process for 10 days (approx. 20 Million steps). Moreover, attempts have been made at developing self-playing agents for Atari games [12], but none of them were able to achieve high performance. More recently, MuZero agent [28] shows that planning can achieve high performance on Atari games but it requires an extensive engineering cost involving a high computational budget. MuZero agent requires more than 2 months of training to train one agent. In this letter, we propose a new gaming Agent called PCTagent. It is arguably the first agent to achieve human-level performance on the Atari games by exhibiting a control architecture similar to living organisms.

1.1 Perceptual Control Theory Modelling

Perceptual Control Theory (PCT) is a computational framework [24–26] to explain and model the behavior of living organisms based on control engineering. Within a PCT system, output (efferent) signals from each control unit within one level of a hierarchy set the goal state for input (afferent) signals at the level below. This allows actions to vary dynamically to control input 'on the fly', so that the planning and learning of actions is unnecessary in many circumstances. A simple, but 'correct', architecture is necessary to achieve a good level of performance. The nervous system of a living organism requires the input functions to construct perceptual signals from

Warren Mansell contributed equally to this work.

✉ Warren Mansell
warren.mansell@curtin.edu.au

Tauseef Gulrez
gtauseef@ieee.org

[1] Department of Computing and Research, Syscon Pty. Ltd., Vanguard Cr., Point Cook, Melbourne 3030, VIC, Australia

[2] School of Population Health, Faculty of Health Sciences, Curtin University, Perth 6102, WA, Australia

the environment, and to organise these with respect to one another [14]. This architecture may have biologically prepared foundations [23], and be further organised through verbal instruction within humans [5]; both of these pathways bypass long periods of training.

A variety of PCT-based computational models have been designed that do not require training, which have emulated naturalistic behaviour within animals [3], human manual tracking [22], human crowd behaviour [17], flyball catching [15] and robotic devices [2, 34]. However, there is little published evidence using established benchmarks within commonly used hardware or modelling environments to compare with approaches that do require training for the learning of actions. This was the aim of the current study.

1.2 Existing AI Methods for Arcade Games

Atari arcade games have been benchmarked primarily using model-free reinforcement learning (RL) algorithms. DQN [18] utilises deep neural network (DNN) to train Q-learning policies by incorporating replay experience and target networks. Several attempts have been made to extend DQN by incorporating bias correction, e.g. DDQN [30], and by prioritising experience replay [27] by architectural modifications [31], and distributional value learning [6]. Some attempts have been made to improve performance by data collection, which increases the cost of environment steps beyond 200 million [1, 19]. Agents developed on proprioceptive inputs [8, 10], model images without using them for planning [21], or combine the benefits of model-based and model-free approaches [13, 20]. Most model-based agents with pixel inputs have thus far been limited to relatively simple control tasks [32].

SimPLe agent [12] learns video frame's pixels and predicts a model in pixelated data format and utilises its predictions to train a proximal policy agent [29]. The model tracks and establishes prediction based on four consecutive frames and incorporates discrete latent variable as an input. The authors evaluate SimPLe on a subset of Atari games for 400k and 2M environment steps, after which the rewards decreased, hence the model over-fitted the environment. As a direct contrast to the above designs, we have used perceptual control theory as a computational framework to show a competitive performance with no training. PCTagent was built on a single GPU Core i7 laptop, for the Atari game environment, within which it outperforms top Atari game agents DreamerV2, DQN, Rainbow [9] and IQN [6] which rest upon years of model-free RL research.

2 Method

2.1 A PCT Agent and Model

PCTagent was developed based on four sub-system hierarchical model of behavior, as shown in Fig. 2 and is available to download under GPL.[1]

The design of a PCTagent is based on a systematic logical analysis of the sensors and effectors of a system, and the performance requirements of the task, that can be specified as an algorithm [7]. Within the Breakout scenario, the sensor can estimate the distance between the paddle (or lever) and ball, which needs to be controlled at zero to perform the task. Yet the effector only commands the rate of button press left or right. Therefore, a hierarchy was constructed that bridges between the control of button press at the bottom level and the control of paddle (lever) - ball distance at the top level. Two intermediate levels were required. The full architecture can therefore be described as follows. The top level of PCTagent controls the visual perception of the distance (D) to be maintained between the paddle and a movable ball in the game environment, as shown in Fig. 1(a)(b). Within Fig. 2, the error (e_1) in the top sub-system sets the reference value (R_2) for the direction control left or right of the paddle (or lever), which is compared to the sensed direction of the paddle (M_D), to generate an error (e_2) that sets the reference value (R_3) for the next level down. The next level down perceives the the position of the paddle (or lever) (P_x) and compares this to the reference value for movement (R_3). Error in this system (e_3) sets the reference value (R_4) for the left or right button press (B_P), and in turn the error of this system (e_4) is transformed to a frequency of button press which determines how far left or right user has to move to hit the ball. An inherent embedded property of button press sub-system is the control of the velocity at which the paddle has to move to catch the ball in case of Pong and Breakout. Within this sub-system, rate of button press is counted by an integrator, which has a limit (currently set to 3), upon reaching max, the integrator resets and reverts to its previous position. Each level of the model contains a k parameter which is the gain of each unit. The gain values for this PCTagent were all set to the value of 1 because this required no a priori assumptions regarding their optimal value.

2.2 ALE Environment and Image Processing

We used the Arcade Learning Environment (ALE) as a platform to empirically assess the performance of our PCTagent. ALE provides an interface for different Atari

[1] https://github.com/PCT-Models/PCTagent_Breakout_Atari

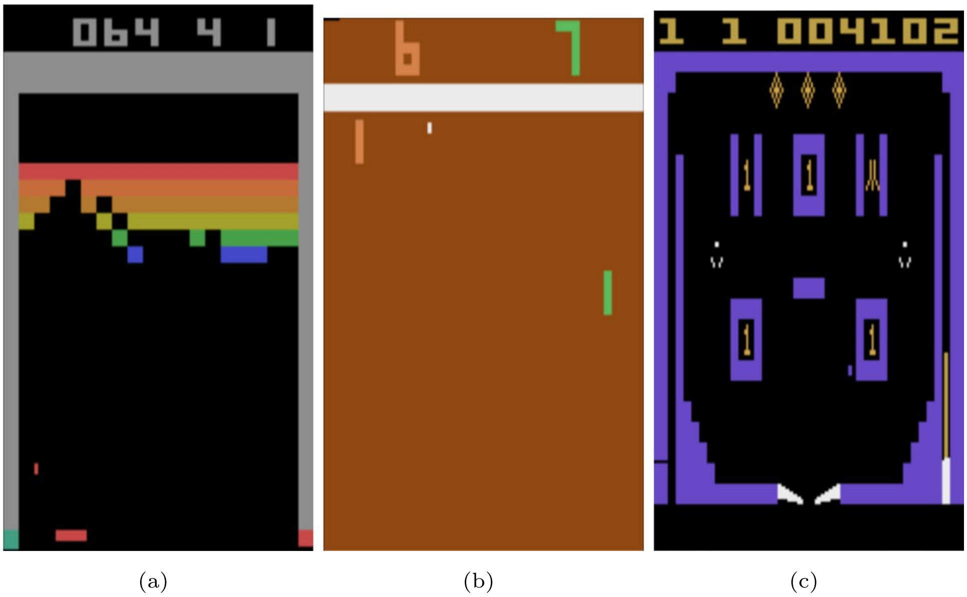


Fig. 1 OpenAI Gym's Atari game environments. (**a**) Breakout (**b**) Pong and (**c**) Video Pinball's RGB frames

2600 game environments which are challenging and engaging for humans. ALE serves as a benchmark environment for the evaluation of AI agents. As an input to our PCTagent we obtained raw Atari frames, which are 210x160 (height,width) pixel image with a 128 colour representation. To make the game computationally efficient, we reduced the input matrix dimensionality by first converting it to binary and then cropping it to obtain a 100x132 region, capturing only the playing area. We do not require any specific square size images, mostly required by 2D machine learning algorithms. The PCTagent was provided with the pre-calculated distance measuring mechanism rather than requiring it to learn from the pixels, hence the tracking of the ball and the paddle were done using simple contour detection function of OpenCV2 module running under Python 3.8.

3 Atari 2600 Games Experiments

First we present our PCTagent's performance results and a comparison of this performance with other state-of-the-art agents as tabulated in Table 1. In Table 1, it can be seen that all of the machine learning game playing agents require millions of frames for training for weeks, whereas since PCTagent does not require any training, hence was ready to play instantly. PCTagent's highest score for Breakout, Pong and Video Pinball was 862, 21 and 203261, respectively.

Atari games e.g. Video Pinball, Breakout and Pong were used to evaluate the performance of our training free PCTagent. The games and PCTagent were run on an Intel Core-i7 single GPU machine with AMD Radeon RX 640 graphics card. All games were tested for different Open Gym environment modes i.e. with/without frame skipping and deterministic mode. The games were ran for 500 episodes and scores data were collected as shown in Fig. 3. In Breakout, a negative reward of -1 for each life lost and in Pong a -1 reward for opponent's successful score was set. It is evident from the results (shown in Fig. 3) that in Breakout it is possible to achieve an average highest score of 800+ and in Pong to win the game with a margin of 5-6 points, at no prior training costs. A significant noise was introduced in the PCTagent when the game were stuck in one position. In the case of Breakout this usually happened in the end when one or two bricks were left to hit, which resulted most of the time in losing a life. In Video Pinball, PCTAgent produced a human like control, i.e. pressing a respective lever to hit the ball. A very realistic high score of 203261 is achieved by PCTAgent. The full development of the agent tackling different modes of the PinBall game was not realised for this stage in the research. Rather, it was just proposed to limit the control to the levers alone. In proposed scenario, a more natural control of levers emerged as opposed to a non-natural way of controlling levers, e.g. indiscriminately vibrating levers regardless at very high frequency. This type of control mechanism may achieve high scores, but it is not consistent with the natural way of playing the game. Hence, PCTAgent is a closer replica of the human way of controlling the levers. But the flip side is that this may not achieve high scores the way that a Deep Learning Agent would control the levers.

4 Conclusion

The idea presented in this Tech Note is to bring this novel (alternative) approach of a learning-less paradigm to the community of robotics, artificial intelligence, automatic control,

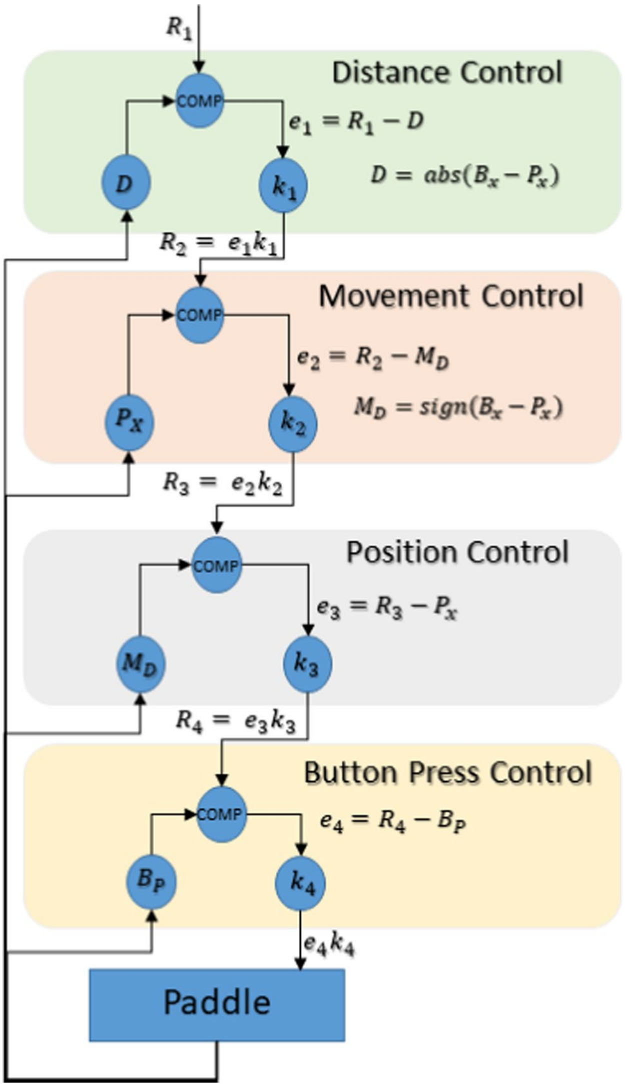
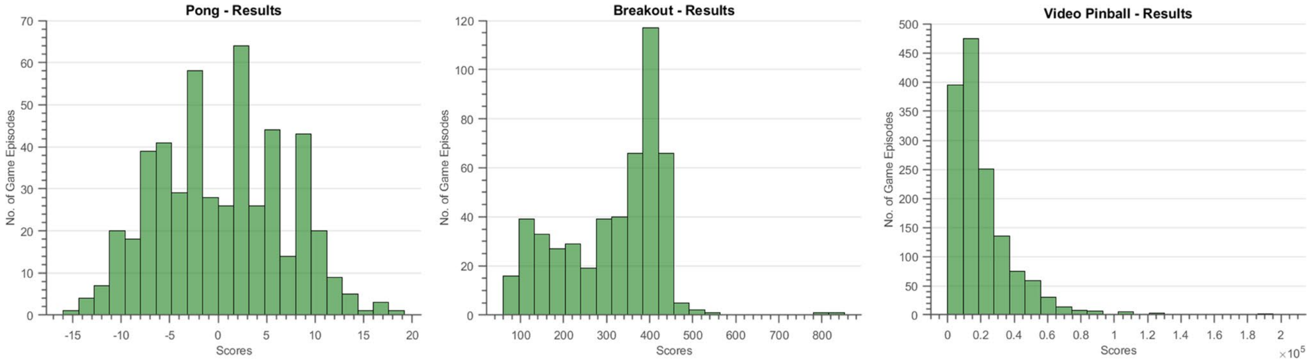


Fig. 2 Hierarchical PCT model for ball and paddle Atari Games

Table 1 Ball and Paddle games results of different ML Agents

Agent Name	Best Score			Frames
	Breakout	Pong	Pinball	(million)
DQN	418.5	21	42684	10
A3C (FF 1 Day)	551.6	11.4	331628.1	100
SimPLe	16.4	5.2	No record	4
RADAR	600+	21	No record	200
Rainbow	120	21	506817.2	200
IQN	734	21	698045	200
DreamerV2	312	20	41860	200
Human World record	864	21	91862206	–
Random	2	-20	16256.9	–
PCTagent (Ours)	862	21	203261	0

machine learning, etc. The Perceptual Control Theoretic (PCT) architecture is used to show that an alternative to reinforcement learning does exist, which requires no training, compared to multiple hours or days of training using very high processing power. We produced one controller based on PCT and tested it, without training, on different Atari 2600 games. This is a first research paper of its kind where a departure from classical way of designing and developing an artificial agent has been proposed. PCT controller performance is human like, naturalistic and matched or exceeded for that of published benchmarks from existing reinforcement and deep learning models. These findings complement earlier studies demonstrating the high performance of PCT controllers within robotics and other areas of research [2, 34]. Indeed, the findings are particularly consistent with an earlier comparison between a PCT controller and an LQR controller for an inverted pendulum robot [11]. This PCT controller required minimal tuning and its performance metrics were superior.



(a) (b) (c)

Fig. 3 Histogram score representation of the 500 episodes of (**a**) Pong (**b**) Breakout and (**c**) Video Pinball games

PCT controllers can provide a robust control solution across environments with no training because they have no need to learn their actions. Instead, they achieve control by varying their outputs on-the-fly to control their inputs by acting against unpredictable disturbances (e.g. obstacles, turbulence, rough ground), unlike reinforcement models. The closed-loop PCT design emulates control systems in nature, which also do not need to learn specific behaviours to operate effectively and efficiently [33]. The choice of input specifications and their hierarchical organisation can be made through a systematic analysis of the sensors, effectors and the task requirements rather than through learning, or inferences made by the researcher [7]. This begs the question of whether reinforcement learning accounts for human skills or whether PCT provides a more accurate model - an architecture of 'priors' proposed to forge future advances in AI [4].

Importantly, PCT controllers can further improve performance through training if required. The learning algorithm specified in PCT - reorganisation - uses random-walk learning to optimise the parameters (e.g. gains) and functions (e.g. specification of inputs that co-vary with target velocity) within a PCT architecture. In particular, we have not looked into sparse rewards problems. However, to draw upon the analogy of unsupervised learning (e.g. Hard Expectation Maximisation) where the probability is either zero (if not one) or one, our approach would be to push the data into the clusters (similar to K-means). We would anticipate that reward shaping as proposed by Maja Mataric's [16] solution towards sparse rewards (reward is either zero or one) could be explored using a PCT architecture and could be explored in the future.

Author Contributions Both authors have contributed equally in this research.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions. There is no funding sought for this research.

Data Availability There is no data and or materials required for this research.

Code Availability The code is available at, https://github.com/PCT-Models/PCTagent_Breakout_Atari.

Declarations

Ethics approval No Ethical approval is required by this research.

Consent for Publication All authors provide consent for publication.

Competing interests All authors agreed with the content and that all gave explicit consent to submit this paper.

Consent to participate No consent form for human participants is required.

References

1. Badia, A.P., Piot, B., Kapturowski, S., et al.: Agent57: Outperforming the atari human benchmark. In: International Conference on Machine Learning, PMLR, pp 507–517 (2020)
2. Barter, J.W., Yin, H.H.: Achieving natural behavior in a robot using neurally inspired hierarchical perceptual control. Iscience **24**(9), 102,948 (2021)
3. Bell, H.C., Bell, G.D., Schank, J.A., et al.: Evolving the tactics of play fighting: Insights from simulating the "keep away game" in rats. Adapt. Behav. **23**(6), 371–380 (2015)
4. Bengio, Y., Lecun, Y., Hinton, G.: Deep learning for AI. Commun. ACM **64**(7), 58–65 (2021)
5. Brown-Ojeda, C., Mansell, W.: Do perceptual instructions lead to enhanced performance relative to behavioral instructions? J. Motor Behav. **50**(3), 312–320 (2018)
6. Dabney, W., Rowland, M., Bellemare, M., et al.: Distributional reinforcement learning with quantile regression. In: Proceedings of the AAAI Conference on Artificial Intelligence (2018)
7. Hawker, B., Moore, R.K.: Robots producing their own hierarchies with dosa; the dependency-oriented structure architect. UK-Robotics and Autonomous Systems (RAS) Network pp 66–68 (2020)
8. Henaff, M., Whitney, W.F., LeCun, Y.: Model-based planning with discrete and continuous actions. arXiv:170507177 (2017)
9. Hessel, M., Modayil, J., Van Hasselt, H., et al.: Rainbow: Combining improvements in deep reinforcement learning. In: Proceedings of the AAAI Conference on Artificial Intelligence (2018)
10. Higuera, J.C.G., Meger, D., Dudek, G.: Synthesizing neural network controllers with probabilistic model-based reinforcement learning. In: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp 2538–2544. IEEE (2018)
11. Johnson, T., Siteng, Z., Cheah, W., et al.: Implementation of a perceptual controller for an inverted pendulum robot. J. Intell. Robot. Syst. **99**(3-4), 683–692 (2020)
12. Kaiser, L., Babaeizadeh, M., Milos, P., et al.: Model-based reinforcement learning for atari. arXiv:190300374 (2019)
13. Kalweit, G., Boedecker, J.: Uncertainty-driven imagination for continuous deep reinforcement learning. In: Conference on Robot Learning, PMLR, pp 195–206 (2017)
14. Marken, R., Kennaway, R., Gulrez, T.: Behavioral illusions: The snark is a boojum. Theory Psychol. **32**(3), 491–514 (2022)
15. Marken, R.S.: Optical trajectories and the informational basis of fly ball catching. J. Exp. Psychol. Hum. Percept. Perform. **31**(3), 340–343 (2005)

16. Mataric, M.J.: Reward functions for accelerated learning. In: Machine learning proceedings 1994, pp 181–189. Elsevier (1994)
17. McPhail, C., Powers, W.T., Tucker, C.W.: Simulating individual and collective action in temporary gatherings. Soc. Sci. Comput. Rev. **10**(1), 1–28 (1992)
18. Mnih, V., Kavukcuoglu, K., Silver, D., et al.: Human-level control through deep reinforcement learning. Nature **518**(7540), 529–533 (2015)
19. Mnih, V., Badia, A.P., Mirza, M., et al.: Asynchronous methods for deep reinforcement learning. In: International conference on machine learning, PMLR, pp 1928–1937 (2016)
20. Nagabandi, A., Kahn, G., Fearing, R.S., et al.: Neural network dynamics for model-based deep reinforcement learning with model-free fine-tuning. In: 2018 IEEE International Conference on Robotics and Automation (ICRA), pp 7559–7566. IEEE (2018)
21. Oh, J., Guo, X., Lee, H., et al.: Action-conditional video prediction using deep networks in atari games. Advances in neural information processing systems 28 (2015)
22. Parker, M.G., Willett, A.B., Tyson, S.F., et al.: A systematic evaluation of the evidence for perceptual control theory in tracking studies. Neurosci. Biobehav. Rev. **112**, 616–633 (2020)
23. Plooij, F.X.: The phylogeny, ontogeny, causation and function of regression periods explained by reorganizations of the hierarchy of perceptual control systems. In: The Interdisciplinary Handbook of Perceptual Control Theory, pp 199–225. Elsevier (2020)
24. Powers, W.T.: Behavior: The control of perception. Aldine Chicago (1973)
25. Powers, W.T.: Living control systems III: The fact of control (2008)
26. Powers, W.T., Clark, R.K., Farland, R.M.: A general feedback theory of human behavior: Part i. Perceptual Motor Skills **11**(1), 71–88 (1960)
27. Schaul, T., Quan, J., Antonoglou, I., et al.: Prioritized experience replay. arXiv:151105952 (2015)
28. Schrittwieser, J., Antonoglou, I., Hubert, T., et al.: Mastering atari, go, chess and shogi by planning with a learned model. Nature **588**(7839), 604–609 (2020)
29. Schulman, J., Wolski, F., Dhariwal, P., et al.: Proximal policy optimization algorithms. arXiv:170706347 (2017)
30. Van Hasselt, H., Guez, A., Silver, D.: Deep reinforcement learning with double q-learning. In: Proceedings of the AAAI Conference on Artificial Intelligence (2016)
31. Wang, Z., Schaul, T., Hessel, M., et al.: Dueling network architectures for deep reinforcement learning. In: International conference on machine learning, PMLR, pp 1995–2003 (2016)
32. Watter, M., Springenberg, J.T., Boedecker, J., et al.: Embed to control: A locally linear latent dynamics model for control from raw images. In: Proceedings of the 28th International Conference on Neural Information Processing Systems, pp 2746–2754 (2015)
33. Yin, H.: The crisis in neuroscience. In: The Interdisciplinary Handbook of Perceptual Control Theory, pp 23–48. Elsevier (2020)
34. Young, R.: A general architecture for robotics systems: A perception-based approach to artificial life. Artif. Life **23**(2), 236–286 (2017)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Tauseef Gulrez received his PhD in robotics and artificial intelligence from Department of Computing, Macquarie University, Sydney, Australia, in September 2008. He held positions of senior research fellow at the University of Salford, Manchester, Biorobotics Lab, University of Washington, Seattle and Cognitive Engineering Lab of Australian Defence Force Academy (ADFA), Canberra. His research interests are in machine learning, robotics and artificial intelligence.

Warren Mansell received his PhD from University of Oxford, and a Doctorate in Clinical Psychology from the Institute of Psychiatry, Kings College, London in 2003. He is currently a Professor at Curtin University, Perth, Australia in Clinical Psychology. The focus of his research is Perceptual Control Theory which he applies across disciplines including mental health, dementia, human action control, communication and robotics.