Reinforcement Learning in Robotics

Patrick M. Pilarski, Ph.D.

Canada CIFAR AI Chair, Dept. of Medicine, University of Alberta
Fellow and Board of Directors, Alberta Machine Intelligence Institute (Amii)
Research Scientist and Edmonton Office Co-Lead, DeepMind
Reinforcement Learning in Biomedical Robotics

Patrick M. Pilarski, Ph.D.

Canada CIFAR AI Chair, Dept. of Medicine, University of Alberta
Fellow and Board of Directors, Alberta Machine Intelligence Institute (Amii)
Research Scientist and Edmonton Office Co-Lead, DeepMind
Techniques from Reinforcement Learning in Bionic Medicine

Patrick M. Pilarski, Ph.D.

Canada CIFAR AI Chair, Dept. of Medicine, University of Alberta
Fellow and Board of Directors, Alberta Machine Intelligence Institute (Amii)
Research Scientist and Edmonton Office Co-Lead, DeepMind
C.O.I. Disclosure

No affiliation (financial or otherwise) with pharmaceutical, medical device or medical communications organizations.

*Other Industry Affiliations:*

Senior Staff Research Scientist and Office Co-Lead, DeepMind

Board of Directors, Alberta Machine Intelligence Institute
950BC - 700BC, The "Cairo Toe" (The University of Manchester),
Video courtesy:
Amii / Chris Onciul
Direct brain-computer interfaces: study participant Jan Scheuermann feeding herself with a robotic limb (University of Pittsburgh / UPMC); http://www.upmc.com/media/media-kit/bci/Pages/default.aspx
cortical implants

bone, muscle, and nerve integration

Brain-body-machine interfaces: “APL’s Modular Prosthetic Limb Reaches New Levels of Operability” (JHU Applied Physics Laboratory); https://youtu.be/-0srXvOQlu0
avatars

e.g.: Avatar startups: https://www.theglobeandmail.com/business/technology/video-ultra-human-like-robots-are-at-the-cutting-edge-of-artificial/
Consumer-Available BCI and BMI
What are some hallmarks of all of these examples?
perception
action
cognition
Extension
Engelbart, 1962
Serino, 2019
Amplification
Ashby, 1956
Tightly Coupled
Licklider, 1960
the control pathway

Hallworth, et al.,
MEC, 2020
machine intelligence

Shehata, et al.,
IEEE Sig. Proc. Magazine, 2021
the feedback path
(mechanical, auditory, visual)
engineered feedback

surgically enhanced feedback

Marasco, et al., Science Robotics, 2021
Hebert, et al., IEEE TNSRE, 2014
machine learned feedback

Parker, et al.,
ICORR, 2019
machine learned bidirectional coordination
To be useful to human users, what properties should we desire of these learning machines?
Tightly coupled interfaces require adaptation and sculpting to individual agents (machine and human) and their unique flow of daily life.
Progress relies on the continual construction of representations, predictions, policies, and models in tightly coupled interfaces.
Main Considerations & Starting Points

Train/test or continual learning?
Continual learning

Pre-trained or tabula rasa?
Minimize prior biases

Relationship or a code channel?
Evolving relationship
One accessible starting point: **Pavlovian control and signalling.**
One accessible starting point: **Pavlovian control** and **signalling**.

Sidebar: I almost always start with prediction learning and ease my way into control or policy learning.
Temporal-Difference Learning Update

Sutton, 1988, Machine Learning
Generalized Value Functions (GVFs)
Pilarski et al., 2012, BioRob

Pavlovian control

Pavlovian control is a process wherein learned, temporally extended predictions

Pavlovian control is a process wherein learned, temporally extended predictions are mapped in a defined way to control actions performed by an agent.

Pavlovian signalling

Butcher et al., 2022; Brenneis et al., 2022; Pilarski et al., 2022.
Pavlovian signalling is a process wherein learned, temporally extended predictions

Butcher et al., 2022; Brenneis et al., 2022; Pilarski et al., 2022.
Pavlovian signalling is a process wherein learned, temporally extended predictions are mapped in a defined way to signals intended for receipt by a decision-making agent.

Butcher et al., 2022; Brenneis et al., 2022; Pilarski et al., 2022.
Pavlovian signalling is a process wherein learned, temporally extended predictions are mapped in a defined way to signals intended for receipt by a decision-making agent, and where these signals are grounded for the sender in the definition of the predictive question and mapping approach that generated them.

Butcher et al., 2022; Brenneis et al., 2022; Pilarski et al., 2022.
The Frost Hollow Experiments

Pavlovian Signalling Co-Agent

Thresholded GVF Prediction

Binary Signal

A/V State

Reward

Event

Action

Human

https://www.youtube.com/watch?v=qdz2wdtkcrk
Brenneis et al., 2022; Butcher et al., 2022; Pilarski et al., 2022.
Continual learning in **motor prediction**.
Parker et al., *IEEE SMC* 2022 (submitted); Parker et al., *ICORR* 2019.

Continual learning in **mode switching**.

Continual learning in **exoskeleton control**.
Faridi et al., *ICORR* 2022.
Continual learning in **motor prediction**. Parker et al., IEEE SMC 2022 (submitted); Parker et al., ICORR 2019.


Continual learning in **exoskeleton control**. Faridi et al., ICORR 2022.
Continual learning in **motor prediction**.
Parker et al., IEEE SMC 2022 (submitted);
Parker et al., ICORR 2019.

Continual learning in **mode switching**.

Continual learning in **exoskeleton control**.
Faridi et al., ICORR 2022.
Continual learning in **motor prediction**.
Parker et al., *IEEE SMC 2022* (submitted); Parker et al., *ICORR 2019*.


Continual learning in **mode switching**.
Edwards et al., *BioRob 2016*.

Continual learning in **exoskeleton control**.
Faridi et al., *ICORR 2022*. 
But also **RL for robot control?**
(Or are you scared of policy learning?)
Examples: 2011-2021

- Identifying patterns with TIDBD
- GVF collections predicting surprise
- LfD from a contralateral limb
- Learned feedback
- Learned joint synergies
- RL policies from human reward
- Pavlovian control in SCI

- Gunther 2020
- Gunther 2018, Pilarski 2016
- Vasan 2017, Vasan 2018
- Parker 2014, 2019
- Pilarski 2013, Sherstan 2015
- Pilarski 2011
- Dalrymple 2020
Examples: 2011-2021

Identifying patterns with TIDBD
GVF collections predicting surprise
LfD from a contralateral limb
Learned feedback
Learned joint synergies
RL policies from human reward
Pavlovian control in SCI

Constructed based on sensorimotor interactions with an individual and what they do, not an objective “task”

Gunther 2020
Gunther 2018, Pilarski 2016
Vasan 2017, Vasan 2018
Parker 2014, 2019
Pilarski 2013, Sherstan 2015
Pilarski 2011
Dalrymple 2020
Situated & Assessable


Video courtesy: Amii / Chris Onciul
Continual learning is important.
RL techniques can be very well suited to continual learning.
Constructing representations, predictions, policies, and models from ongoing experience lets tightly coupled interfaces align & specialize to individual human (or machine) agents and needs.
Pavlovian control and signalling is a natural gateway to more complex continual interactions.
machine learned bidirectional coordination
Continually learning tightly coupled intelligent systems
Thank you and questions!

Jacqueline Hebert
Richard Sutton
Craig Chapman
Albert Vette
Vivian Mushahwar
Adam White
Joseph Modayil
Jason Carey
Mahdi Tavakoli
Kim Adams
Martin Ferguson-Pell
Simon Grange
Liping Qi
Matt Botvinick
Todd Murphey
K. Ming Chan
Erik Scheme
Michael Bowling
Kory Mathewson
Craig Sherstan
Elnaz Davoodi
Thomas Degris
Michael Johanson
Ahmed Shehata
Johannes Gunther
Florian Strub
Ivana Kajic

Claudio Castellini
Jon Sensinger
Paul Marasco
Aida Valevicius
Hiroki Tanikawa
Michael Rory Dawson
Mayank Rehani
Glyn Murgatroyd
Dylan Brenneis
Andrew Butcher
Leslie Acker
Andrew Bolt
Adam Parker
Heather Williams
Ola Kalinowska
Alden Christianson
Ann Edwards
Alex Kearney
Nadia Ady
Ewen Lavoie
Katherine Schoepp
Pouria Faridi
Travis Dick
Vivek Veeriah
Riley Dawson

Quinn Boser
Jaden Travnik
Gautham Vasan
Anna Koop
Kodi Cheng
Emma Durocher
Devin Bradburn
Helen Zhao
Liam Jack
Roshan Shariff
Nathan Wisniski
Ben Hallworth

... and all the other members of our teams and labs advising or contributing to aspects the presented work.

With additional funding and support from: NSERC, CFI, Canada Research Chairs, Canada CIFAR AI Chairs, DARPA HAPTIX, GRHF, UHF, Alberta Innovates, and the Government of Alberta.