



Prediction, Knowledge, and Explainability: Examining the Use of General Value Functions in Machine Knowledge

Alex Kearney^{1*}, Johannes Günther^{1,2} and Patrick M. Pilarski^{1,2,3}

¹ Reinforcement Learning and Artificial Intelligence Lab, Department of Computing Science, University of Alberta, Edmonton, AB, Canada, ² Alberta Machine Intelligence Institute, Edmonton, AB, Canada, ³ Department of Medicine, University of Alberta, Edmonton, AB, Canada

Within computational reinforcement learning, a growing body of work seeks to express an agent's knowledge of its world through large collections of predictions. While systems that encode predictions as General Value Functions (GVFs) have seen numerous developments in both theory and application, whether such approaches are explainable is unexplored. In this perspective piece, we explore GVFs as a form of explainable AI. To do so, we articulate a subjective agent-centric approach to explainability in sequential decision-making tasks. We propose that prior to explaining its decisions to others, a self-supervised agent must be able to introspectively explain decisions to itself. To clarify this point, we review prior applications of GVFs that involve human-agent collaboration. In doing so, we demonstrate that by making their subjective explanations public, predictive knowledge agents can improve the clarity of their operation in collaborative tasks.

Keywords: reinforcement learning, General Value Functions, knowledge, prediction, explainability

OPEN ACCESS

Edited by:

Volker Steuber,
University of Hertfordshire, United Kingdom

Reviewed by:

Kristen Wolff,
Chalmers University of Technology, Sweden

*Correspondence:

Alex Kearney
hi@alexkearney.com

Specialty section:

This article was submitted to
AI in Business,
a section of the journal
Frontiers in Artificial Intelligence

Received: 01 December 2021

Accepted: 10 March 2022

Published: 31 March 2022

Citation:

Kearney A, Günther J and Pilarski PM
(2022) Prediction, Knowledge, and
Explainability: Examining the Use of
General Value Functions in Machine
Knowledge.
Front. Artif. Intell. 5:826724.
doi: 10.3389/frai.2022.826724

1. A SUBJECTIVE VIEW ON AGENTS

It is advantageous for agents to learn about the world that they are operating in. An agent deployed for a sufficient amount of time will face circumstances unforeseen by its creator, making the ability adapt and react to new experiences essential. Computational reinforcement learning (RL) methods that adapt based on experience have had numerous recent successes in both research domains and applied settings (Mnih et al., 2015; Silver et al., 2017; Lazic et al., 2018; Bellemare et al., 2020); however, how agents make decisions is often opaque, leading to concerns about how agents will perform in new circumstances (Adadi and Berrada, 2018; Holzinger, 2018). In this article, we concern ourselves with how artificial agents in sequential decision-making problems can explain their decisions. In particular, we focus on systems that model their world by learning large collections of predictions encoded as General Value Functions (GVFs) (White et al., 2015), an area of Machine Intelligence in which explainability has yet to be explored. To do so, we argue for a subjective, and self-supervised approach to explainability.

2. PREDICTIONS: THE FOUNDATION OF AGENT PERCEPTION

Evidence suggests that the best ways for an agent to construct an independent conception of the world is through their subjective experiences (Brooks, 2019). One way for agents to learn about

their experience is to model their world as many interrelated predictions (or *forecasts*) (White et al., 2015). These predictions can be as simple as “If I keep walking forwards, will I bump into something?”, and can be so complex as to map out the entirety of the environment an agent exists in Ring (2021).

By learning predictions, an agent’s conception of the world can be composed exclusively of interrelations of sensations and actions over time. Forecasts can be made and learned in an entirely self-supervised fashion: agents do not require labeled examples nor a human ontology to form abstract relationships. This enables agents to construct new abstract relationships from its own stream of experience by adding new predictions over time (Sherstan et al., 2018b; Veeriah et al., 2019; Kearney et al., 2021) without human prompting or input.

Understanding perception through the lens of prediction has a basis not only in AI, but also in biological systems, such as humans (Wolpert et al., 1995; Rao and Ballard, 1999). Humans are constantly making predictions about what they expect to see next in order to anticipate and react to their environment (Gilbert, 2009; Clark, 2015). It is no surprise then, that predictive approaches to machine perception have also been successful. Recent examples include industrial laser welding (Günther, 2018), robot navigation (Daftry et al., 2016), facilitate walking in decerebrate cats (Dalrymple et al., 2020), and control of bionic upper-limb prosthetics (Edwards et al., 2016a). A core theme that unifies this broad collection of research is (1) the ability to flexibly adapt to new situations by predicting future stimuli, and (2) learning predictions using experience.

There have been numerous advances in the algorithms underpinning predictive machine perception, and applications in real-world settings (e.g., Sherstan et al., 2015; Edwards et al., 2016a; Günther, 2018; Veeriah et al., 2019; Dalrymple et al., 2020; Schlegel et al., 2021); however, explainability of predictive knowledge systems is as of yet unexplored. In this article, we take a first look at explainability in General Value Function (GVF) systems by reflecting on recent advances in the literature. In doing so, we highlight the importance of introspection: How an agent can examine its own subjective experience and explain decisions to itself. We then focus on how the private subjective explanation can be made public, and explore examples where such explanations led to success in joint-action systems between humans and agents.

3. VALUE FUNCTIONS AND HOW THEY ARE ESTIMATED

To explore the explainability of predictive knowledge systems, we must first understand how they are specified and learned (Figure 1). The core component of predictive knowledge systems are their learned GVFs: forecasts of future *returns* of signals from the agent’s environment. One of the best understood ways of making temporally-extended predictions is through temporal difference (TD) learning (Sutton, 1988). In temporal difference learning, time is broken down into discrete *time-steps* $t_0 \dots t_n$. On each time-step, the agent observes the environment as o_t : a tuple of values that include everything the agent can perceive (e.g.,

sensor readings, inputs from a camera). Using TD learning the *value* of a future signal $V(o; \pi, \gamma, c) = \mathbb{E}[G_t | o_t = o]$ is estimated, given the agent’s experience.

General Value Functions (GVFs) learn the value specified by three *question parameters*: a *cumulant* c , a *policy* π , and a discount γ . The cumulant is the signal of interest being accumulated. The policy describes the behavior the prediction is conditioned on. The discount function describes the temporal horizon over which we wish to make the prediction and how the cumulant is accumulated. Using these three parameters, functions about the agent’s subjective experience can be expressed as *returns* $G_t^\pi = \sum_{k=0}^{\infty} \gamma^k C_{t+k+1}$. An agent can then learn a model that estimates the *return*, or *value* of a GVF given its current sensations *via* TD learning (Algorithm 1).

4. EXPLANATION TO SELF AND OTHERS: HOW AN AGENT INTERPRETS ITS OWN ESTIMATES AND SHARES THEM

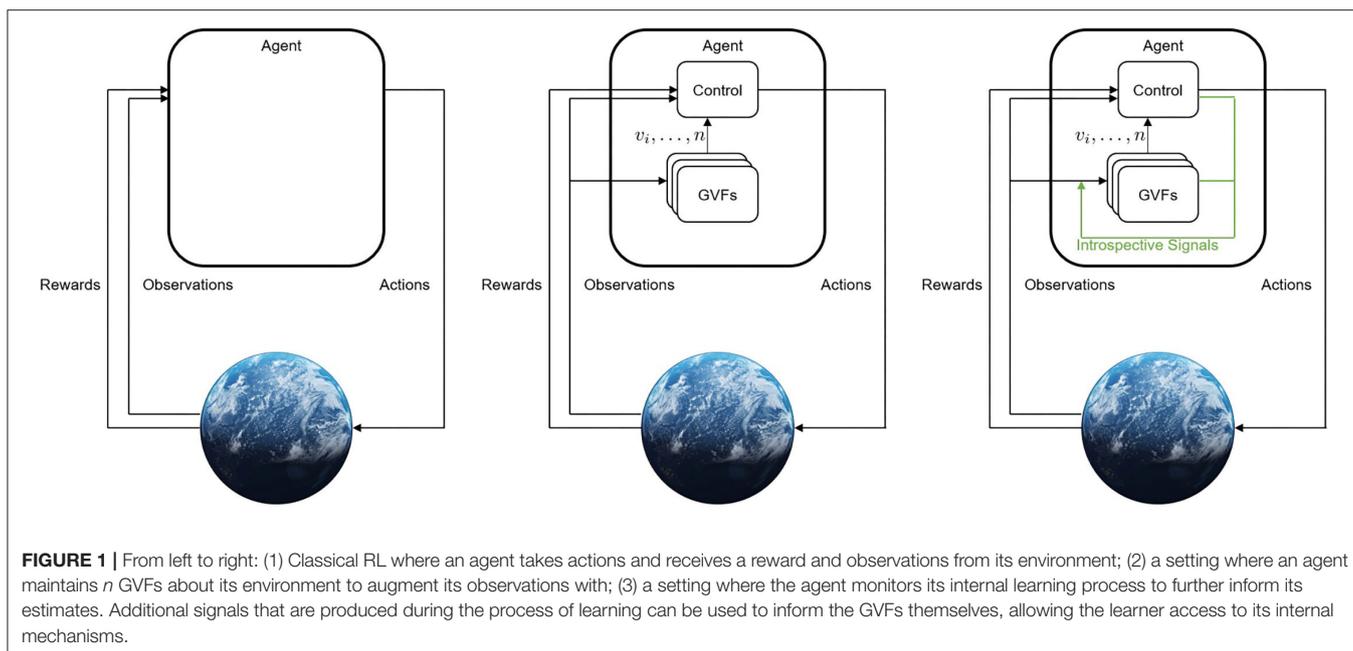
If the question parameters of a GVF are explainable, then it is apparent what a prediction is about by looking at the question parameters themselves. The cumulant $c(o_t)$, policy $\pi(o_t, a_t)$, and discount $\gamma(o_t, a_t, o_{t+1})$ describe what a prediction is about in terms of environmental stimuli to the agent. While these parameters do not provide an intuitive explanation in terms of natural language or symbols, the cumulant, discount, and policy are inherently interpretable.

While we can determine *what* a prediction is about by examining the question parameters, the question remains if we determine *why* a decision or estimate was made by looking at the model that produced the estimate: e.g., given an observation o_t , why is the value $v(o_t)$? Whether we can determine why an estimate was made depends on whether function approximation is used.

If a particular GVF is a linear combination of features and weights, it is inherently interpretable (Molnar, 2020; Ghassemi et al., 2021): coefficients w_i are directly multiplied with each input o_i , and each coefficient w_i determines how an estimate $v(o_t)$ was reached by weighting the observations. However, many value functions cannot be characterized by a linear relationship. By using *function approximation* (e.g., Sherstov and Stone, 2005; Mnih et al., 2015; Travník and Pilarski, 2017) non-linearity is added to the inputs, producing a new vector $x(o_t)$ with which we

Algorithm 1 | TD(λ).

- 1: Initialize eligibility traces $z \in 0^n$, weights $w \in 0^n$; initialize a step-size $0 < \alpha \leq 1$ and eligibility decay $0 \leq \lambda \leq 1$; observe environment o
 - 2: Repeat for each observation o and cumulant c :
 - 3: $\delta \leftarrow c + \gamma w^\top x(o_{t+1}) - w^\top x(o_t)$
 - 4: For $i = 1, 2, \dots, n$:
 - 5: $z_i \leftarrow z_i \gamma \lambda + x_i(o_t)$
 - 6: $w_i \leftarrow w_i + \alpha_i \delta z_i$
 - 7: $o_t \leftarrow o_{t+1}$
-



linearly combine with the models weights in order to produce an estimate $v(o_t) = w^\top x(o_t)$. The weights w are no longer inherently interpretable since they are learned such that they combine with $x(o_t)$.

Interpreting the model's weights when using function approximation and making sense of its estimates is challenging, as is the case with non-linear many ML methods. However, an agent may use *post-hoc* interpretation (Molnar, 2020) methods to justify its decisions. In this discussion, we contribute a distinction between two forms of explanation: *introspective explanation*—how an agent analyses and evaluates its own internal signals and decisions; and, *exteroceptive explanation*—how an agent communicates its internal signals to other agents or individuals.

5. INTROSPECTIVE EXPLANATION

The problem of introspective explanation within the context of predictive knowledge is then the problem of an agent deciding whether or not an estimate—whether a particular GVF within the agent's model—is sufficient to rely on for decision-making. Prior to rationalizing a decision to external agents and human designers, an agent must commit to a decision itself—an agent must choose an action a_t using the information available to itself o_t . In predictive knowledge systems, GVF forecasts are often a key component of this observation o_t (Edwards et al., 2016b; Günther, 2018; Kearney et al., 2021; Ring, 2021; Schlegel et al., 2021). However, not all predictions are created equally.

The accuracy and precision of predictions are dependent on a variety of factors, including the learning methods used to generate the model, the hyperparameters that modulate learning, and the inputs or features available to learn a model from. Two models of the environment can have different numerical values, despite being *about* the same thing. It is important to

differentiate between accurate and inaccurate estimates when they are used as inputs for decision-making tasks. Poor inputs can lead to catastrophic errors; moreover, these errors may not be immediately apparent from the agent's (or human designer's) perspective (Kearney et al., 2020).

Looking only at immediate estimates $v(o_t)$, faults in an agent's own model and GVF estimates are inscrutable. On each time-step the agent is aware of its estimate $v(o_t)$ and the *bootstrapped* value observed $c + \gamma v(o_{t+1})$. Although error δ can be calculated from these values (line 6, **Algorithm 1**) to further learn the estimates, error is often not an input in o_t that influences an agent's decision-making. Without additional information, an agent has no means to discern how reliable an estimate in its model is.

5.1. Estimating Variance: A Self-Assessment of Certainty

By monitoring the variance of estimates (Sherstan et al., 2018a), an agent can take into account both the model estimates and the certainty in those estimates during decision-making (Sherstan et al., 2016). By evaluating the qualities of its own estimates and learning progress, an agent can develop a sense of confidence: “how much do I trust that my estimates are accurate?” This ability to incorporate *post-hoc* interpretations (Molnar, 2020) of its own model in decision-making, is a form of self-explanation. In this case, an agent chooses an action not only because of predicted values, but the relative certainty in those values.

An estimate from a newly initialized GVF will often be poor, and for good reason: it has not had time to refine its estimates through experience. An agent which examines the certainty of estimates alone cannot differentiate between newly added components of its model, and genuinely poor estimates. It is possible to use summary statistics of internal learning to capture measurements of learning progress (Kearney et al., 2019;

Linke, 2021). By monitoring how much it has learned about its environment, an agent can use this information as an auxiliary reward to shape how it moves about its environment to further learn useful behaviors (Linke, 2021). In this case actions are taken based on not only what is believed to be rewarding, but also what action would enable the most learning.

5.2. Unexpected Error: As Self-Assessment of Surprise

Unexpected Demon Error (UDE) is a metric designed to express an agent's "surprise" about a sensation (White et al., 2015). By calculating UDE, an agent can monitor unexpected changes in either the environment or its own physical body (Pilarski and Sherstan, 2016; Günther et al., 2018; Wong, 2021). Neither error encountered while learning, nor error due to random noise in the environment factor into UDE. As regular predictable sources of error are filtered out, the UDE represents salient sources of error that would not be otherwise noticeable to an agent.

5.3. Feature Relevance: A Self-Assessment of Explanatory Variables

Learning processes are governed by parameters that dictate how learning occurs. One important parameter is the *step-size* or *learning-rate*. The step-size α determines how much the model's parameters w are updated every time-step based on the observed error δ . The value of a GVF's step-size α can be set by a meta-learning process (Günther et al., 2019; Kearney et al., 2019). Not only do these adaptive step-sizes improve learning, they can also provide useful information about the operation of an agent (Günther et al., 2019). Learned step-size values can also detect anomalies in the agent's observation stream, including those that indicate a hardware or sensor failure of the agent (Günther et al., 2020).

Each internal measurement we discussed is a fragment of an agent's inner life that can directly be accessed and monitored by the agent itself. By introspecting and monitoring their own internal signals, an agent can augment their observations: an agent can act on not only its model of the environment, but its interpretation of how the model was constructed. By building up the private, subjective experience of an agent, explanation is no longer grounded in symbols that are pre-defined by human designers, but is rooted in the agent's own conception of the world. Each internal measurement we discuss is a measurement that depends only on what is available to the agent. In this case, the agent's ability to explain is limited only by the functions of internal and external signals it can monitor. As an agent adds new predictions to learn, introspective measurements can be added to assess each new GVF of the agent's model, without human intervention. Each individual GVF can be learned separately from the others, and thus each GVF that makes up our model can be interpreted independent of the rest. It is not necessary to assess a collection of predictions monolithically, as is the case with many neural-network based world models. By prioritizing agents that introspect, we provide a basis for explanation that begins with the agent explaining to itself.

6. MAKING INTROSPECTION PUBLIC: EXPLANATION TO OTHERS

Having given an agent the ability to explain its beliefs and decisions to itself, we can begin to make the agent's private inner life public: we can make overt the subjective view of the agent. We view exterospective explanation through two lenses: (1) the behavior the agent undertakes, and (2) the estimates the agent learns and produces. In this section, we show that making the internal signals of the agent's inner life external is a viable form of communication and explanation.

6.1. Reflexes: Communication *via* Fixed Responses

The simplest way an agent can communicate with its outside world is by taking an action. The simplest action an agent can take is a reflex: an automatic fixed response to some external stimuli. Reflexes are a foundational form of behavior in both complex and simple biological systems (Gomezano and Kehoe, 1975; Jirenhed et al., 2007; Pavlov, 2010). Artificial agents can learn to anticipate signals (Ludvig et al., 2012), and perform reflexively similar to biological systems (Modayil and Sutton, 2014). Reflexes are inherently explainable, because they are deterministic. By reflexively acting, an agent is making external its internal estimate for a given stimuli. While deterministic, differentiating between reflexes and the stochastic behavior of regular decision-making is not always clear. Moreover, when behaviors are not fixed, interpreting the cause for a particular behavior is more challenging: many different stimuli could possibly elicit the same response from the agent.

6.2. Predictions: Communication of Inner Estimates

The challenge of interpreting agents based on behavior alone is evident in existing work using GVF-based systems on biomedical devices and assistive devices (Pilarski et al., 2013; Edwards et al., 2016a,b). In these applications, humans interact with machines that build behaviors based on predictions learned from the machine's subjective experience (Pilarski et al., 2013; Sherstan et al., 2015; Edwards et al., 2016a). Applications of predictive knowledge systems where humans and agents collaborate to control a system, have been shown to improved performance over non-adaptive systems an individual controls directly (Edwards et al., 2016b); however, one overriding theme is that human participants have consistently expressed a lack of trust in these systems where internal predictions that led to decisions were not visible to them (c.f., Edwards et al., 2016b). The actions of the agent in the environment are an insufficient form of communication with human collaborators. When the internal predictions made by a system were made visible in their raw form to a human partner, participants were able to both understand and fruitfully make decisions in the shared environment (Edwards et al., 2016b). In a task where a human collaborator was in direct control of

a robotic third arm, when predictions of motor torque (a proxy for impact in a workspace) were mapped to vibration on a human collaborator's arm, the user was better able to perform precise positioning tasks (Parker et al., 2019). Similarly, by using audio cues to communicate a machine's internal prediction of the value of objects in a Virtual Reality environment, humans were able to perform better in a foraging task (Pilarski et al., 2019). In a task where an artificial agent had control over some aspects of a bionic limb, external communication of the predictions driving control were needed to enable well-timed collaboration with a jointly acting human (Edwards et al., 2016b). Put simply—humans often find it uncomfortable to collaborate with a learning machine on a task if they do not have a window into the predictions that inform decision making. By making the agent's internal estimates public, humans are able to learn about the agent and build trust through joint experience. Exploring how internal signals can be made public provides a promising frontier for future work exploring joint-action GVF systems, and a ideal setting to put into practice the suggestions presented in this perspective article.

7. DISCUSSION

In this article, we explore GVFs within the context of explainable AI. We discuss how the question and answer parameters that make up a GVF can be interpreted by human designers. We emphasize that self-supervised agents benefit from introspection, arguing that introspection is a form of self-explanation. We explore a wide range of collaborative human-agent GVF systems, highlighting how these agents' success is dependent on making the internal, introspective signals public to their human collaborators. Collaborative human-agent systems are

more effective when an agent's internal estimates are made public to collaborators, even though such signals are not necessarily in human terms or symbols. Although predictive knowledge systems are a relatively new area of research, they are a promising way of developing self-supervised and explainable AI.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

AK led the research and analysis for this manuscript. JG provided the figures. JG and PP supervised this research. PP secured funding for this research. All authors contributed to the writing, editing, and approval of this manuscript for submission.

FUNDING

This research was undertaken, in part, thanks to funding from the Canada Research Chairs program, the Canada CIFAR AI Chairs program, the Canada Foundation for Innovation, the Alberta Machine Intelligence Institute, Alberta Innovates, and the Natural Sciences and Engineering Research Council. AK was supported by scholarships and awards from NSERC.

ACKNOWLEDGMENTS

The authors thank Anna Koop and Dylan Brenneis for providing feedback on an early draft of this manuscript.

REFERENCES

- Adadi, A., and Berrada, M. (2018). Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE Access* 6, 52138–52160. doi: 10.1109/ACCESS.2018.2870052
- Bellemare, M. G., Candido, S., Castro, P. S., Gong, J., Machado, M. C., Moitra, S., et al. (2020). Autonomous navigation of stratospheric balloons using reinforcement learning. *Nature* 588, 77–82. doi: 10.1038/s41586-020-2939-8
- Brooks, R. (2019). *A Better Lesson*. Robots, AI, and other stuff. Available online at: <https://rodnebrooks.com/a-better-lesson> (accessed December 12, 2019).
- Clark, A. (2015). *Surfing Uncertainty: Prediction, Action, and The Embodied Mind*. Oxford University Press.
- Daftry, S., Zeng, S., Bagnell, J. A., and Hebert, M. (2016). "Introspective perception: learning to predict failures in vision systems," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (Daejeon: IEEE)*, 1743–1750. doi: 10.1109/IROS.2016.7759279
- Dalrymple, A. N., Roszko, D. A., Sutton, R. S., and Mushahwar, V. K. (2020). Pavlovian control of intraspinal microstimulation to produce over-ground walking. *J. Neural Eng.* 17, 036002. doi: 10.1088/1741-2552/ab8e8e
- Edwards, A. L., Dawson, M. R., Hebert, J. S., Sherstan, C., Sutton, R. S., Chan, K. M., et al. (2016a). Application of real-time machine learning to myoelectric prosthesis control: A case series in adaptive switching. *Prosthet. Orthot. Int.* 40, 573–581. doi: 10.1177/0309364615605373
- Edwards, A. L., Hebert, J. S., and Pilarski, P. M. (2016b). "Machine learning and unlearning to autonomously switch between the functions of a myoelectric arm," in *2016 6th IEEE International Conference on Biomedical Robotics and Biomechanics (BioRob) (IEEE)*, 514–521. doi: 10.1109/BIOROB.2016.7523678
- Ghassemi, M., Oakden-Rayner, L., and Beam, A. L. (2021). The false hope of current approaches to explainable artificial intelligence in health care. *Lancet Digit. Health* 3, e745?e750. doi: 10.1016/S2589-7500(21)00208-9
- Gilbert, D. (2009). *Stumbling on Happiness*. New York, NY: Vintage Canada.
- Ormezeano, I., and Kehoe, E. (1975). Classical conditioning: some methodological-conceptual issues. *Handb. Learn. Cogn. Process.* 2, 143–179.
- Günther, J. (2018). *Machine intelligence for adaptable closed loop and open loop production engineering systems* (Ph.D. thesis). Technische Universität München, Munich, Germany.
- Günther, J., Ady, N. M., Kearney, A., Dawson, M. R., and Pilarski, P. M. (2020). Examining the use of temporal-difference incremental delta-bar-delta for real-world predictive knowledge architectures. *Front. Robot. AI* 7, 34. doi: 10.3389/frobt.2020.00034
- Günther, J., Kearney, A., Ady, N. M., Dawson, M. R., and Pilarski, P. M. (2019). "Meta-learning for predictive knowledge architectures: a case study using TIDBD on a sensor-rich robotic arm," in *Proceedings of the 18th International Conference on Autonomous Agents and Multi Agent Systems (Montreal, QC)*, 1967–1969.
- Günther, J., Kearney, A., Dawson, M. R., Sherstan, C., and Pilarski, P. M. (2018). "Predictions, surprise, and predictions of surprise in general value function architectures," in *AAAI 2018 Fall Symposium on Reasoning and Learning in Real-World Systems for Long-Term Autonomy (Arlington, VA)*, 22–29.

- Holzinger, A. (2018). "From machine learning to explainable AI," in *2018 World Symposium on Digital Intelligence for Systems and Machines* (Kosice: IEEE), 55–66. doi: 10.1109/DISA.2018.8490530
- Jirehned, D.-A., Bengtsson, F., and Hessler, G. (2007). Acquisition, extinction, and reacquisition of a cerebellar cortical memory trace. *J. Neurosci.* 27, 2493–2502. doi: 10.1523/JNEUROSCI.4202-06.2007
- Kearney, A., Koop, A., Günther, J., and Pilarski, P. M. (2021). Finding useful predictions by meta-gradient descent to improve decision-making. *arXiv [Preprint]*. arXiv: 2111.11212. doi: 10.48550/arXiv.2111.11212
- Kearney, A., Koop, A., and Pilarski, P. M. (2020). What's a good prediction? Issues in evaluating general value functions through error. *arXiv [Preprint]*. arXiv: 2001.08823. doi: 10.48550/arXiv.2001.08823
- Kearney, A., Veeriah, V., Travník, J., Pilarski, P. M., and Sutton, R. S. (2019). Learning feature relevance through step size adaptation in temporal-difference learning. *arXiv preprint arXiv:1903.03252*. doi: 10.48550/arXiv.1903.03252
- Lazic, N., Boutilier, C., Lu, T., Wong, E., Roy, B., Ryu, M., et al. (2018). Data center cooling using model-predictive control. *Adv. Neural Inform. Process. Syst.* 31, 3814–3823.
- Linke, C. (2021). *Adapting Behavior via Intrinsic Reward* (MSc thesis). University of Alberta, Edmonton, United States. Available online at: https://era.library.ualberta.ca/items/3a8b07d6-bab6-4916-8cd1-2485e84309a0/view/e04326c5-4e76-4128-a410-c290c2d2f43d/Linke_Cameron_202102_MSc.pdf
- Ludvig, E. A., Sutton, R. S., and Kehoe, E. J. (2012). Evaluating the TD model of classical conditioning. *Learn. Behav.* 40, 305–319. doi: 10.3758/s13420-012-0082-6
- Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., et al. (2015). Human-level control through deep reinforcement learning. *Nature* 518, 529–533. doi: 10.1038/nature14236
- Modayil, J., and Sutton, R. S. (2014). "Prediction driven behavior: learning predictions that drive fixed responses," in *Workshops at the Twenty-Eighth AAAI Conference on Artificial Intelligence* (Quebec City, QC).
- Molnar, C. (2020). *Interpretable Machine Learning*. Available online at: <https://christophm.github.io/interpretable-ml-book/>
- Parker, A. S., Edwards, A. L., and Pilarski, P. M. (2019). "Exploring the impact of machine-learned predictions on feedback from an artificial limb," in *2019 IEEE 16th International Conference on Rehabilitation Robotics* (Toronto, ON: IEEE), 1239–1246. doi: 10.1109/ICORR.2019.8779424
- Pavlov, P. I. (2010). Conditioned reflexes: an investigation of the physiological activity of the cerebral cortex. *Ann. Neurosci.* 17, 136. doi: 10.5214/ans.0972-7531.1017309
- Pilarski, P. M., Butcher, A., Johanson, M., Botvinick, M. M., Bolt, A., and Parker, A. S. (2019). "Learned human-agent decision-making, communication and joint action in a virtual reality environment," in *4th Multidisciplinary Conference on Reinforcement Learning and Decision Making* (Ann Arbor, MI), 302–306.
- Pilarski, P. M., Dawson, M. R., Degris, T., Carey, J. P., Chan, K. M., Hebert, J. S., et al. (2013). Adaptive artificial limbs: a real-time approach to prediction and anticipation. *IEEE Robot. Automat. Mag.* 20, 53–64. doi: 10.1109/MRA.2012.2229948
- Pilarski, P. M., and Sherstan, C. (2016). "Steps toward knowledgeable neuroprostheses," in *2016 6th IEEE International Conference on Biomedical Robotics and Biomechanics (BioRob)* (University Town: IEEE), 220–220. doi: 10.1109/BIOROB.2016.7523626
- Rao, R. P., and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* 2, 79–87. doi: 10.1038/4580
- Ring, M. B. (2021). Representing knowledge as predictions (and state as knowledge). *arXiv [Preprint]*. arXiv: 2112.06336. doi: 10.48550/arXiv.2112.06336
- Schlegel, M., Jacobsen, A., Abbas, Z., Patterson, A., White, A., and White, M. (2021). General value function networks. *J. Artif. Intell. Res.* 70, 497–543. doi: 10.1613/jair.1.12105
- Sherstan, C., Bennett, B., Young, K., Ashley, D. R., White, A., White, M., et al. (2018a). "Directly estimating the variance of the -return using temporal difference methods," in *Conference on Uncertainty in Artificial Intelligence* (Monteray, CA), 63–72.
- Sherstan, C., Machado, M. C., and Pilarski, P. M. (2018b). "Accelerating learning in constructive predictive frameworks with the successor representation," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (Madrid: IEEE), 2997–3003. doi: 10.1109/IROS.2018.8594242
- Sherstan, C., Modayil, J., and Pilarski, P. M. (2015). "A collaborative approach to the simultaneous multi-joint control of a prosthetic arm," in *2015 IEEE International Conference on Rehabilitation Robotics (ICORR)* (University Town: IEEE), 13–18. doi: 10.1109/ICORR.2015.7281168
- Sherstan, C., White, A., Machado, M. C., and Pilarski, P. M. (2016). "Introspective agents: confidence measures for general value functions," in *International Conference on Artificial General Intelligence* (New York: Springer), 258–261. doi: 10.1007/978-3-319-41649-6_26
- Sherstov, A. A., and Stone, P. (2005). "Function approximation via tile coding: automating parameter choice," in *International Symposium on Abstraction, Reformulation, and Approximation* (Airth Castle: Springer), 194–205. doi: 10.1007/11527862_14
- Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., et al. (2017). Mastering the game of go without human knowledge. *Nature* 550, 354–359. doi: 10.1038/nature24270
- Sutton, R. S. (1988). Learning to predict by the methods of temporal differences. *Mach. Learn.* 3, 9–44. doi: 10.1007/BF00115009
- Travník, J. B., and Pilarski, P. M. (2017). "Representing high-dimensional data to intelligent prostheses and other wearable assistive robots: a first comparison of tile coding and selective Kanerva coding," in *Proceedings of the International Conference on Rehabilitation Robotics* (London: IEEE), 1443–1450. doi: 10.1109/ICORR.2017.8009451
- Veeriah, V., Hessel, M., Xu, Z., Rajendran, J., Lewis, R. L., Oh, J., et al. (2019). Discovery of useful questions as auxiliary tasks. *arXiv [Preprint]*. arXiv: 1909.04607. Available online at: <https://arxiv.org/pdf/1909.04607.pdf>
- White, A. (2015). *Developing a Predictive Approach to Knowledge* (Ph. D. thesis). University of Alberta, Edmonton, United States. Available online at: https://era.library.ualberta.ca/items/7f973a64-35c9-4109-9a79-d87edb44ae52/view/ece8302b-75f6-472c-93cb-2e88f322b06f/White_Adam_M_201506_PhD.pdf
- Wolpert, D. M., Ghahramani, Z., and Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science* 269, 1880–1882. doi: 10.1126/science.7569931
- Wong, A. (2021). *Intelligent machine reliability with general value functions* (MSc thesis). University of Alberta, Edmonton, United States.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Kearney, Günther and Pilarski. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.