Open Issues in Open World Learning

WALTER J. SCHEIRER
UNIVERSITY OF NOTRE DAME

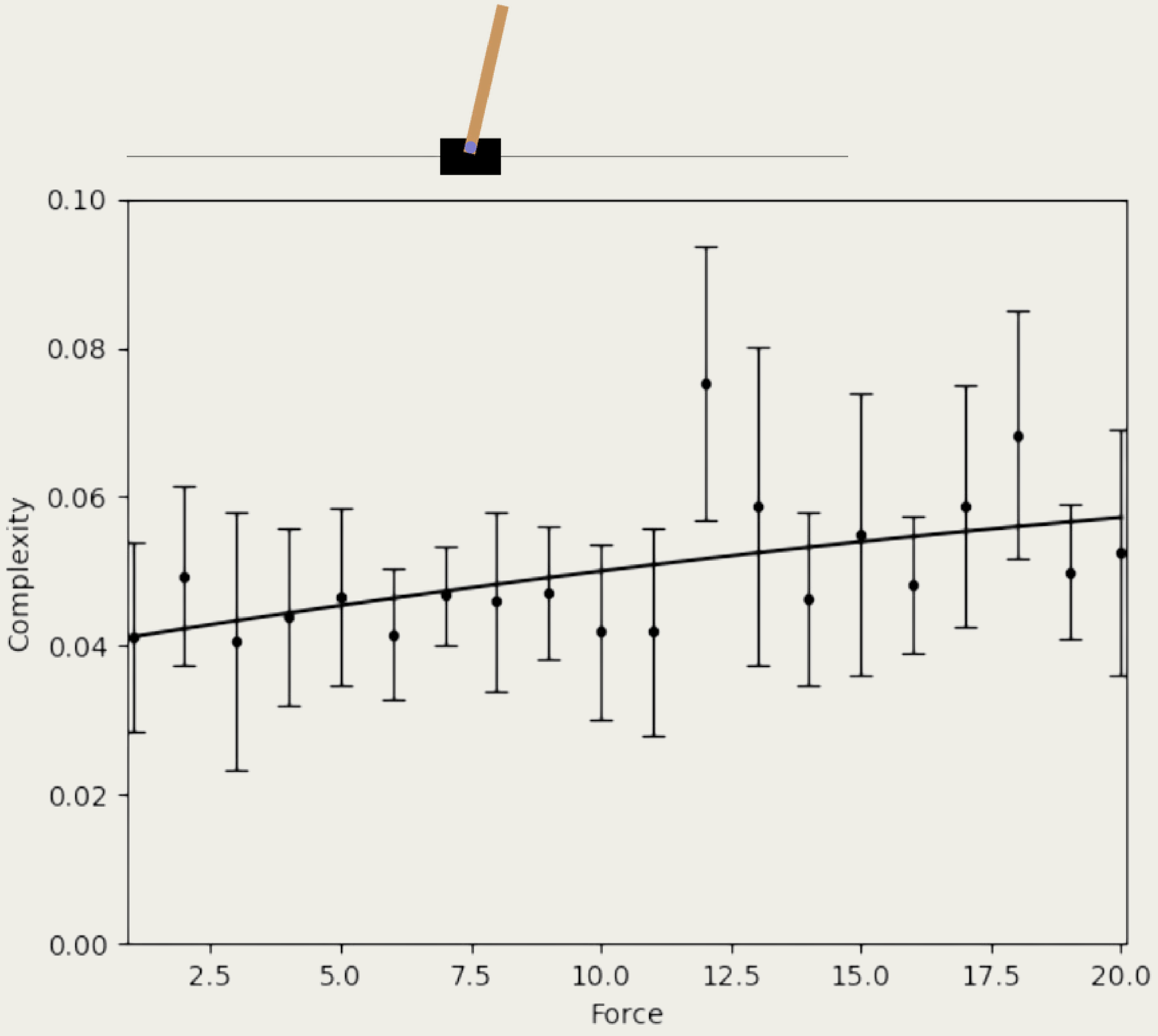
WHY DON'T WE HAVE SAFE SELF-DRIVING CARS IN 2024?

Significant shortcoming of the big data strategy for machine learning: it necessarily **under-samples** the world, often drastically.

“There are ‘countably infinite things’ that can possibly happen” [on the road].

- Neil Lawrence on X (2019)

CARTPOLE



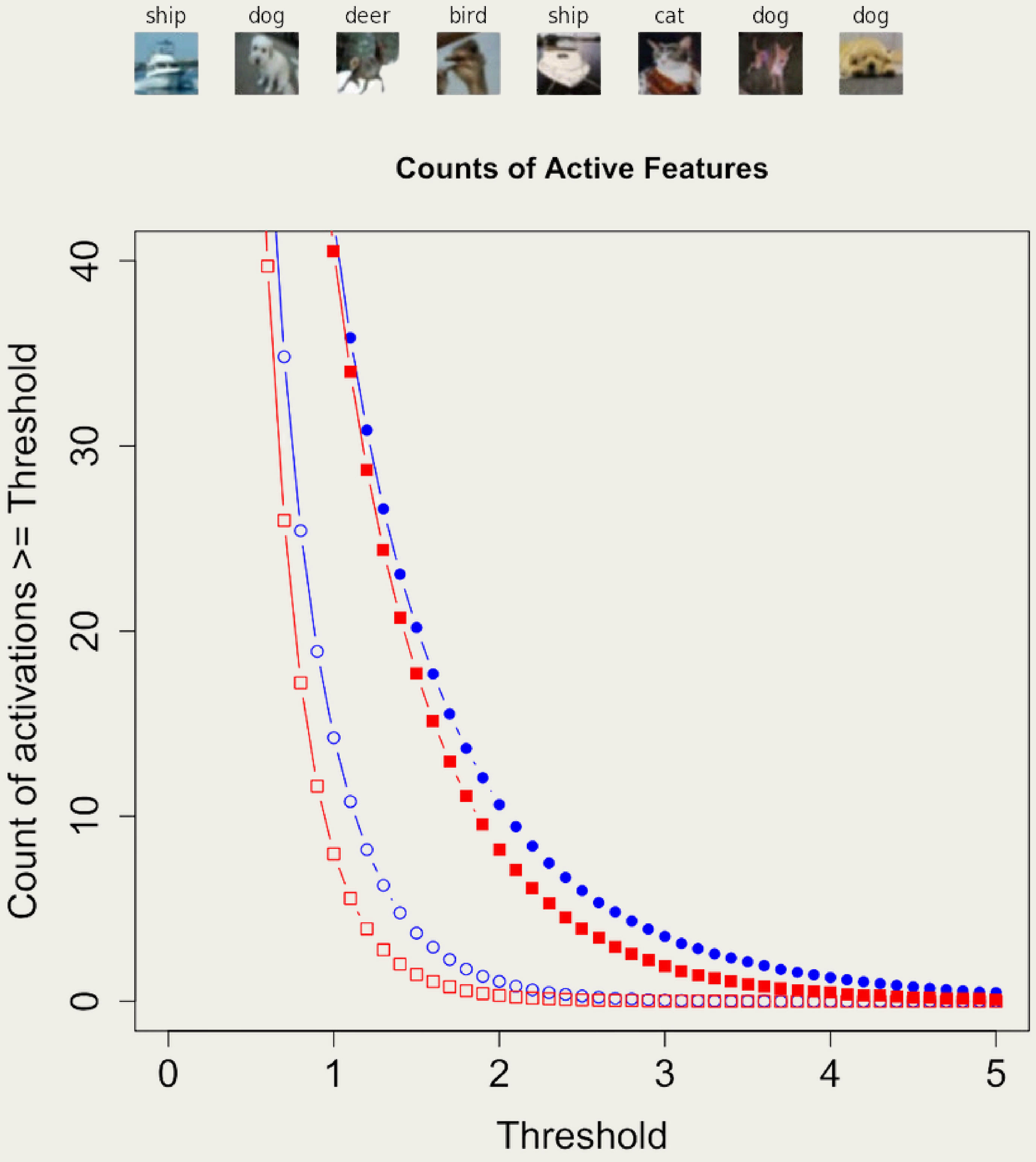
domain score policy

$$Complexity(\mu) = \int_{V\ \min_\mu}^{V\ \max_\mu} V_\mu^{MDL(\pi)}\, dV$$

min. description length

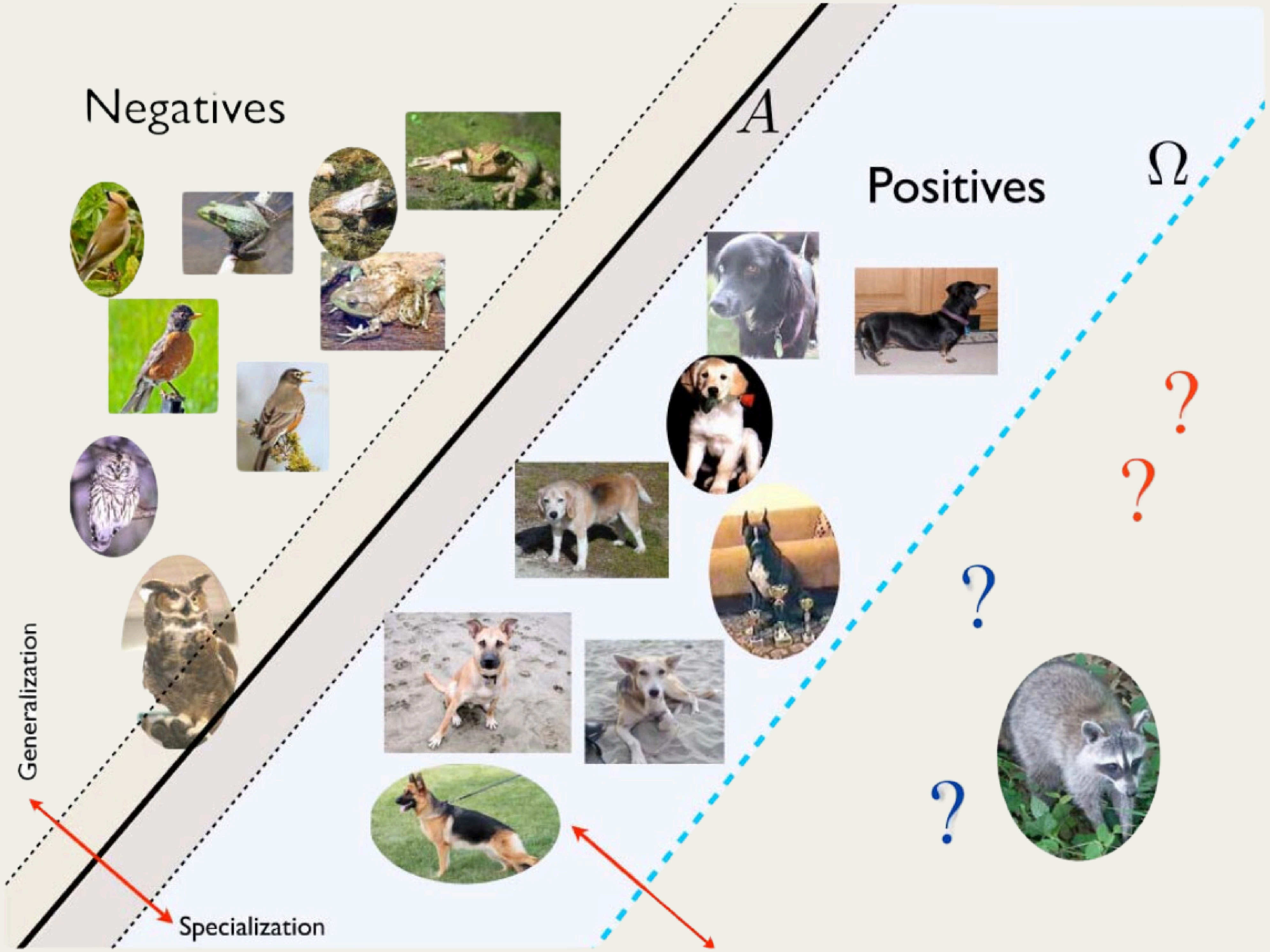
Pereyda and Holder, “Measuring the Complexity of Domains Used to Evaluate AI Systems, ”AAAI Spring Symposium on Designing AI for Open Worlds 2022

CIFAR-10



Dietterich and Guyer, “The Familiarity Hypthesis: Explaining the Behavior of Deep Open Set Methods,” *Pattern Recognition* 2022

WHY IS NOVELTY SUCH A CONFOUND?



Training

We always have incomplete knowledge of the world at training time for open world domains.

Basis

There is no basis from which to generalize to novel data from known data, both of which are typically far away from each other in a feature space.

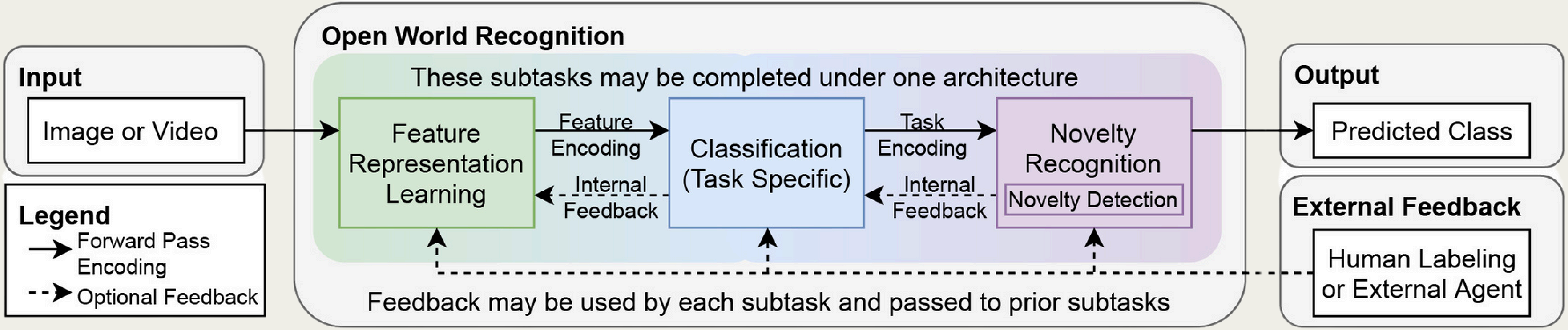
Generation

New things appear in an open environment all of the time. This isn't restricted to object classes. Novel activities and interactions are also significant confounds.

Scheirer et al., "Toward Open Set Recognition," *IEEE T-PAMI* 2012


NOVELTY NEEDS TO BE MANAGED
WITH REINFORCEMENT LEARNING, RIGHT?
WITH REINFORCEMENT LEARNING, RIGHT?
imgflip.com

OPEN WORLD AGENT TEMPLATE



Note 1: Novelty detection and characterization turn out to be really hard

Note 2: Efficient incremental learning is necessary

Note 3: Classifier research not very popular right now

Prijatelj et al., “Human Activity Recognition in an Open World,” arXiv 2022

DARPA SAIL-ON PROGRAM

DARPA DEFENSE ADVANCED RESEARCH PROJECTS AGENCY

EXPLORE BY TAG

ABOUT US / OUR RESEARCH / NEWS / EVENTS / WORK WITH US /

› Defense Advanced Research Projects Agency › Our Research › Science of Artificial Intelligence and Learning for Open-world Novelty

Science of Artificial Intelligence and Learning for Open-world Novelty (SAIL-ON) (Archived)

Current artificial intelligence (AI) systems excel at tasks defined by rigid rules – such as mastering the board games Go and chess with proficiency surpassing world-class human players. However, AI systems aren't very good at adapting to constantly changing conditions commonly faced by troops in the real world – from reacting to an adversary's surprise actions, to fluctuating weather, to operating in unfamiliar terrain. For AI systems to effectively partner with humans across a spectrum of military applications, intelligent machines need to graduate from closed-world problem solving within confined boundaries to open-world challenges characterized by fluid and novel situations.

The Science of Artificial Intelligence and Learning for Open-world Novelty (SAIL-ON) program intends to research and develop the underlying scientific principles, general engineering techniques, and algorithms needed to create AI systems that act appropriately and effectively in novel situations that occur in open worlds. The program's goals are to develop scientific principles to quantify and characterize novelty in open-world domains, create AI systems that react to novelty in those domains, and demonstrate and evaluate these systems in a selected DoD domain.

https://www.darpa.mil/program/science-of-artificial-intelligence-and-learning-for-open-world-novelty

DARPA SAIL-ON PROGRAM: ORGANIZATION

Open World Novelty Hierarchy			
Single Entities	**Phase 1**	**1**	**Objects:** New classes, attributes, or representations of non-volitional entities.
	Phase 2	**2**	**Agents:** New classes, attributes, or representations of volitional entities.
		3	**Actions:** New classes, attributes, or representations of external agent behavior.
Multiple Entities		**4**	**Relations:** New classes, attributes, or representations of static properties of the relationships between multiple entities.
		5	**Interactions:** New classes, attributes, or representations of dynamic properties of behaviors impacting multiple entities.
Complex Phenomena	**Phase 3**	**6**	**Rules:** New classes, attributes, or representations of global constraints that impact all entities.
		7	**Goals:** New classes, attributes, or representations of external agent objectives.
		8	**Events:** New classes, attributes, or representations of series of state changes that are not the direct result of volitional action by an external agent or the SAIL-ON agent.

DARPA SAIL-ON PROGRAM: METRICS

Type	Name	Measure	Definition
Detection (Distribution Change Detection)	M1	$\overline{\mathrm{FN}}_{\mathrm{CDT}}$	Mean # of FNs among CDTs
	M2	CDT%	% of CDTs (among all Trials)
	M2.1	FP%	% of Trials with at least 1 FP
Accommodation (Task Performance)	M3,M4	NRP	$\frac{\sum P_{Post,\alpha}}{\sum P_{Pre,\beta}}$
	AM1	Overall PTI (OPTI)	$\frac{\sum P_{Post,\alpha}}{\sum P_{Post,\alpha} + \sum P_{Post,\beta}}$
	AM2	Asymptotic PTI (APTI)	$\frac{\sum_{i=N_T-m}^{N_T} P_{Post,\alpha}}{\sum_{i=N_T-m}^{N_T} P_{Post,\alpha} + \sum_{i=N_T-m}^{N_T} P_{Post,\beta}}$

DARPA SAIL-ON PROGRAM: WHAT DID WE LEARN?

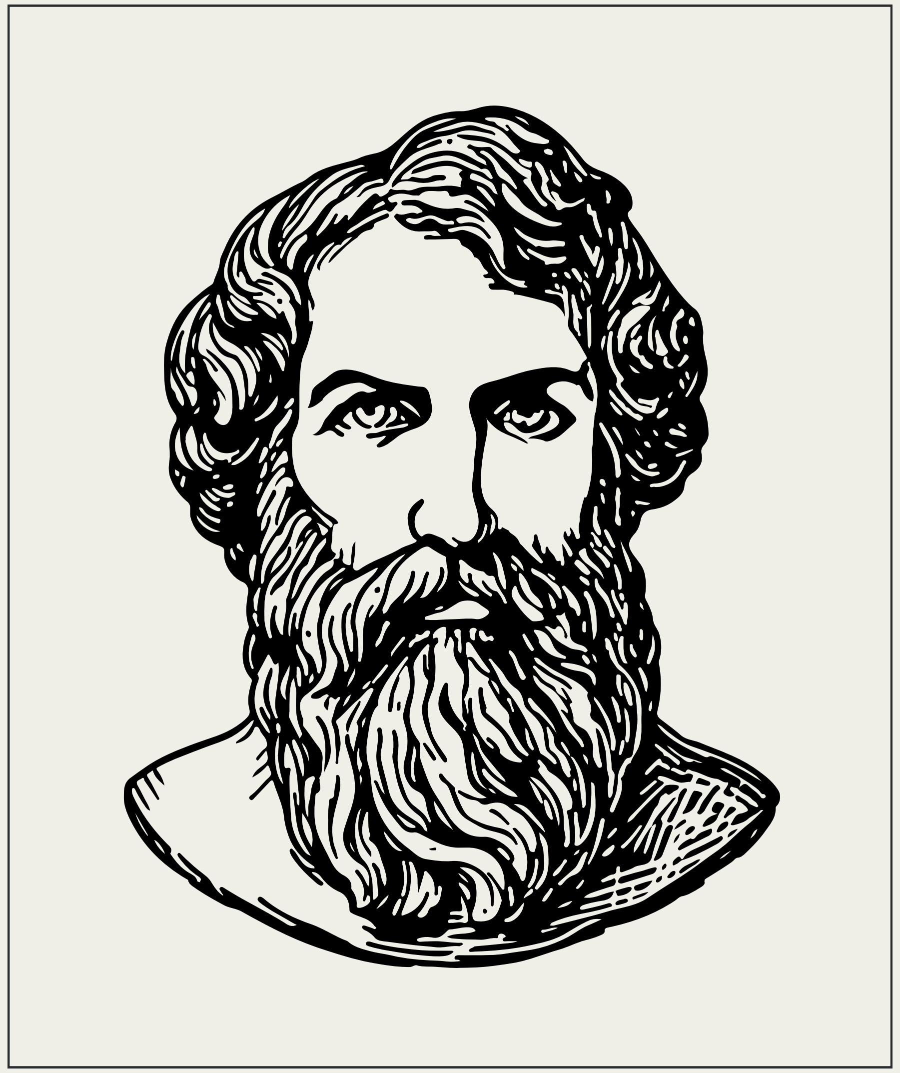
Open Issues in Open World Learning	Novelty Category
1. Need for Better Developed Theories of Novelty	Theory of Novelty
2. Differences Between Activity and Perceptual Domains	Theory of Novelty
3. Domain Independence	Design of Agents
4. Better Representation for Novelty Learning	Design of Agents
5. Robustness to Novelty Versus Novelty Detection and Characterization	Design of Agents
6. Risk-Based Reasoning	Design of Agents
7. Spectrum of Partial Knowledge the System Designer Has About Novelties	Evaluation of Agents
8. Lack of Measures Specific to Open World Learning	Evaluation of Agents

1. NEED FOR BETTER DEVELOPED THEORIES OF NOVELTY

Plato (*Timaeus*): new things are generated by reconfiguring existing material into a new form through the guidance of set patterns.

Kuhn: novelty reduces to the perception of things in an environment that are new to the observer.[1]

Langley: environmental change in a generative mode is the basis of novelty. [2]

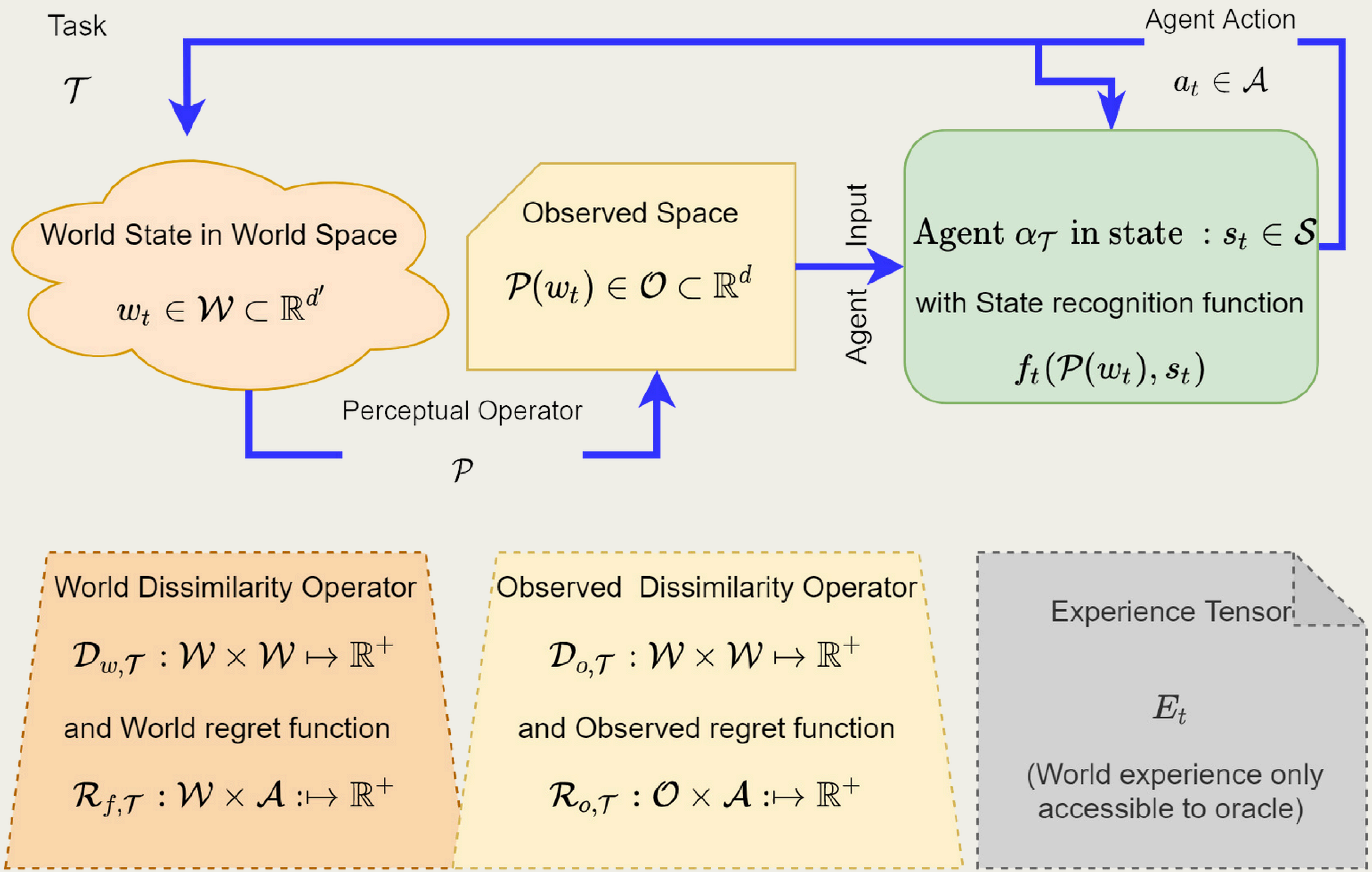


Theory of Novelty

1. Kuhn, "Second thoughts on paradigms," *The Structure of Scientific Theories* 2, 1974, pp. 459–482.

2. Langley, "Open-world learning for radically autonomous agents," AAAI 2020.

1. NEED FOR BETTER DEVELOPED THEORIES OF NOVELTY

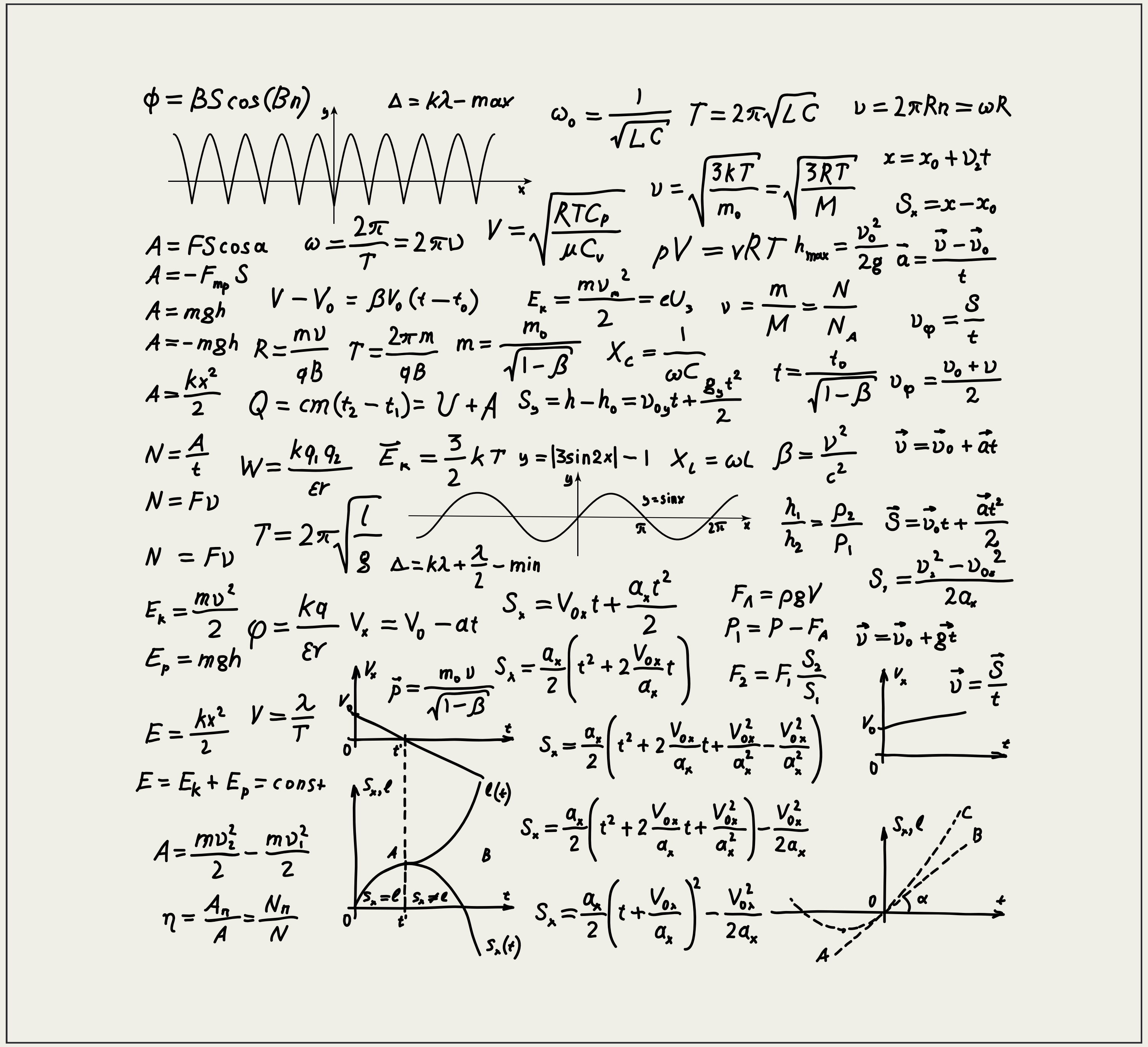


Boult et al., “Towards a unifying framework for formal theories of novelty,“ AAAI 2021

1. NEED FOR BETTER DEVELOPED THEORIES OF NOVELTY

Open Questions:

- Given the two overarching framings of environmental novelty and agent-centric novelty, is it possible to reconcile them into a single theory?
- Is there any theoretical basis for a novelty hierachy?
- Does any extant theory help make predictions about agents and their interactions with the environment that can help guide agent design?



2. DIFFERENCES BETWEEN ACTIVITY AND PERCEPTUAL DOMAINS

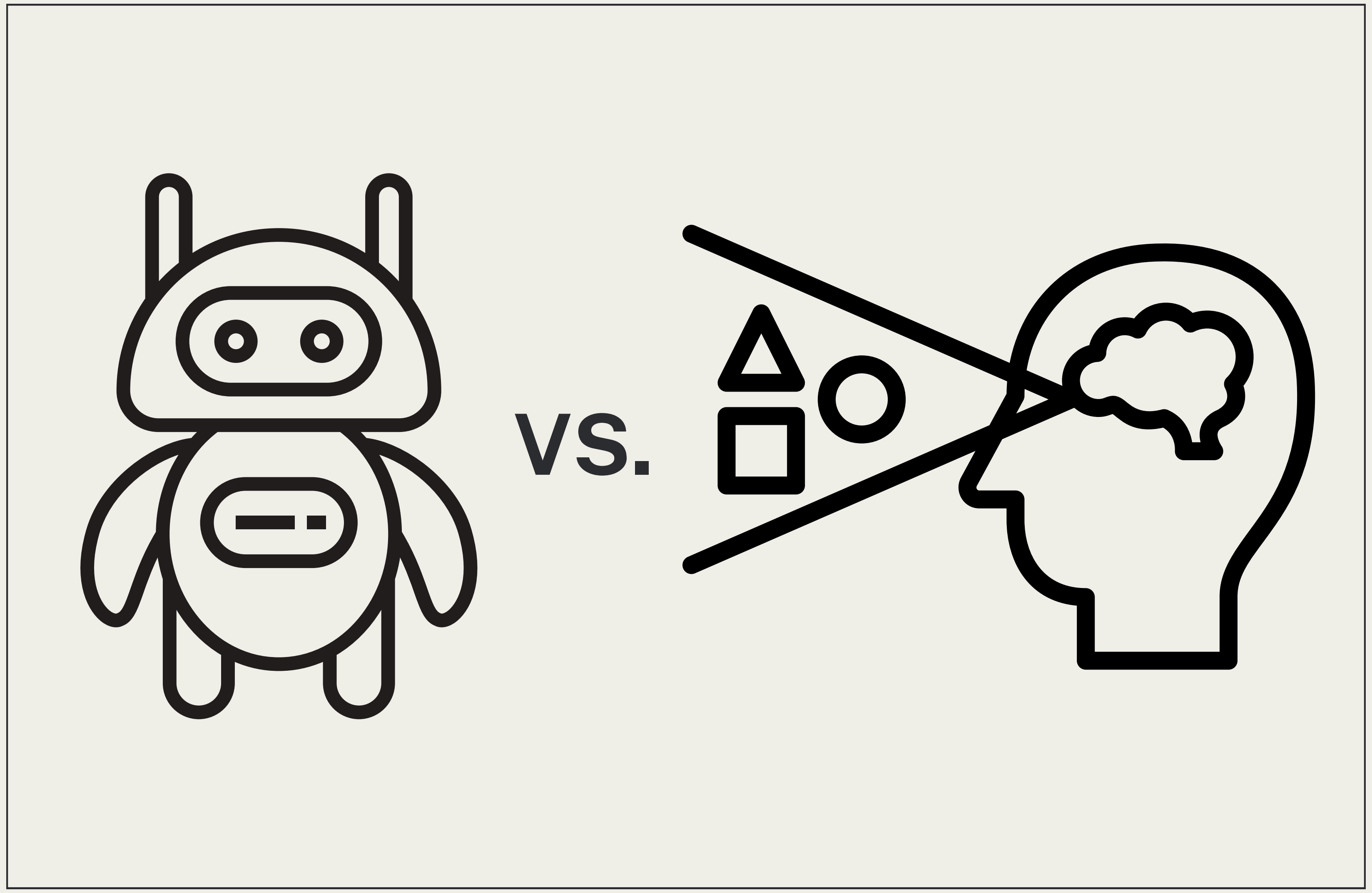
Activity Domains: interactive environments where an agent attempts to achieve and objective by making state transitions that are favorable to it.

Example: Robot agent in the physical world.

Perceptual Domains: focus on the sensing aspect of agents in a non-interactive environment.

Example: Object recognition in computer vision.

Theory of Novelty

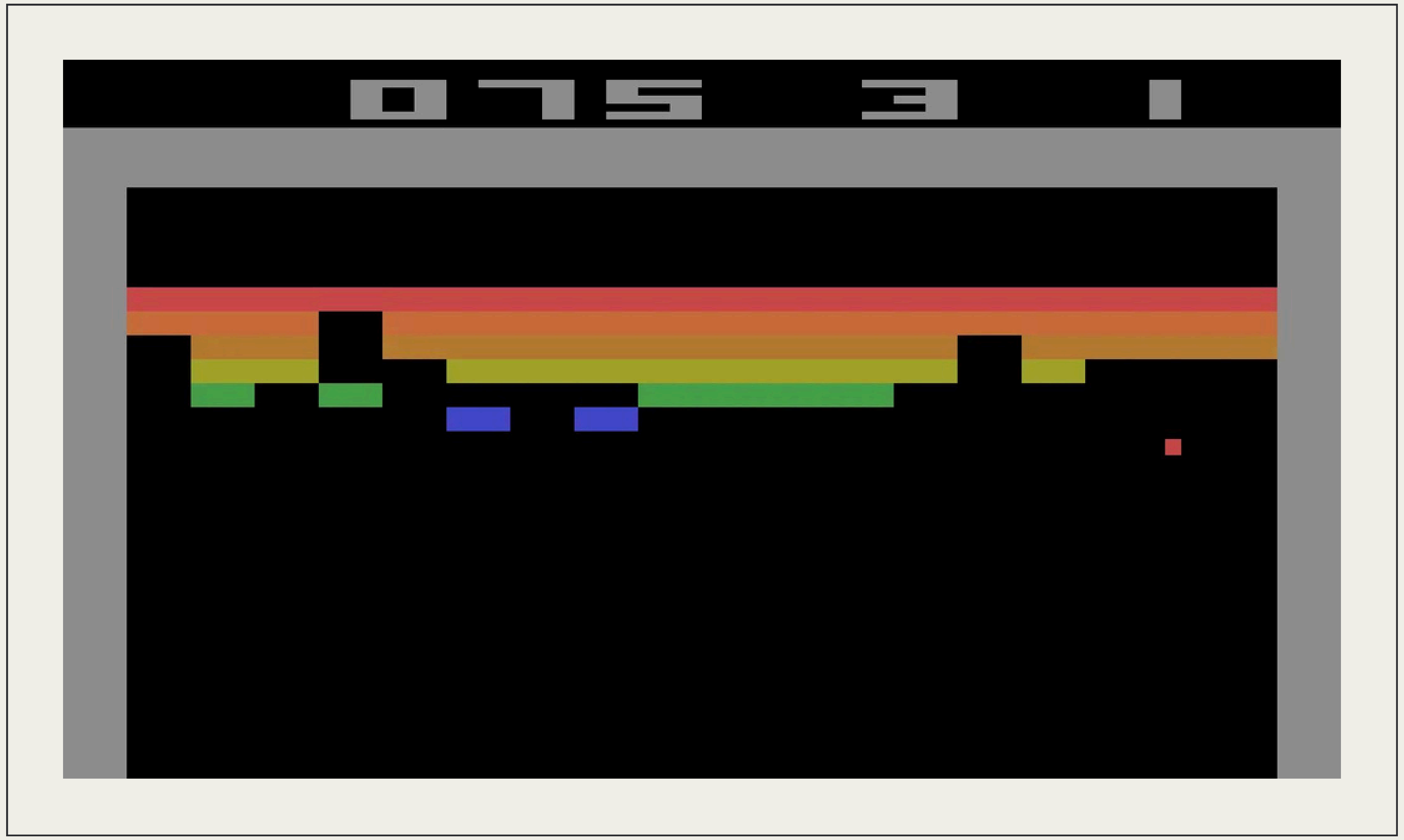


2. DIFFERENCES BETWEEN ACTIVITY AND PERCEPTUAL DOMAINS

Can the two be reconciled?

Some hints that this is possible:

- Mobile robotics
- DeepMind-style video game play (Mnih et al. NeurIPS 2013)
- Predictive Coding (Burachas et al. AAAI Spring Symposium on Designing AI for Open Worlds 2022)
- Dissimilarity assessment through agent observation (Boult et al. AAAI 2021)



Mnih et al., “Playing Atari with deep reinforcement learning,” NeurIPS 2013

Burachas et al., “Metacognitive mechanisms for novelty processing: Lessons for AI,” AAAI Spring Symposium on Designing AI for Open Worlds 2022

Boult et al., “Towards a unifying framework for formal theories of novelty,“ AAAI 2021

3. DOMAIN INDEPENDENCE

Bold claims about deep learning generalization have never been true: deep nets are limited to just a single domain – the one associated with the training data (Chollet 2017).

Planning suffers from a similar problem: if an agent moves to a new environment, the old plan may provide it no useful information.

Design of agents

Chollet, “The limitations of deep learning,” The Keras Blog, 2017

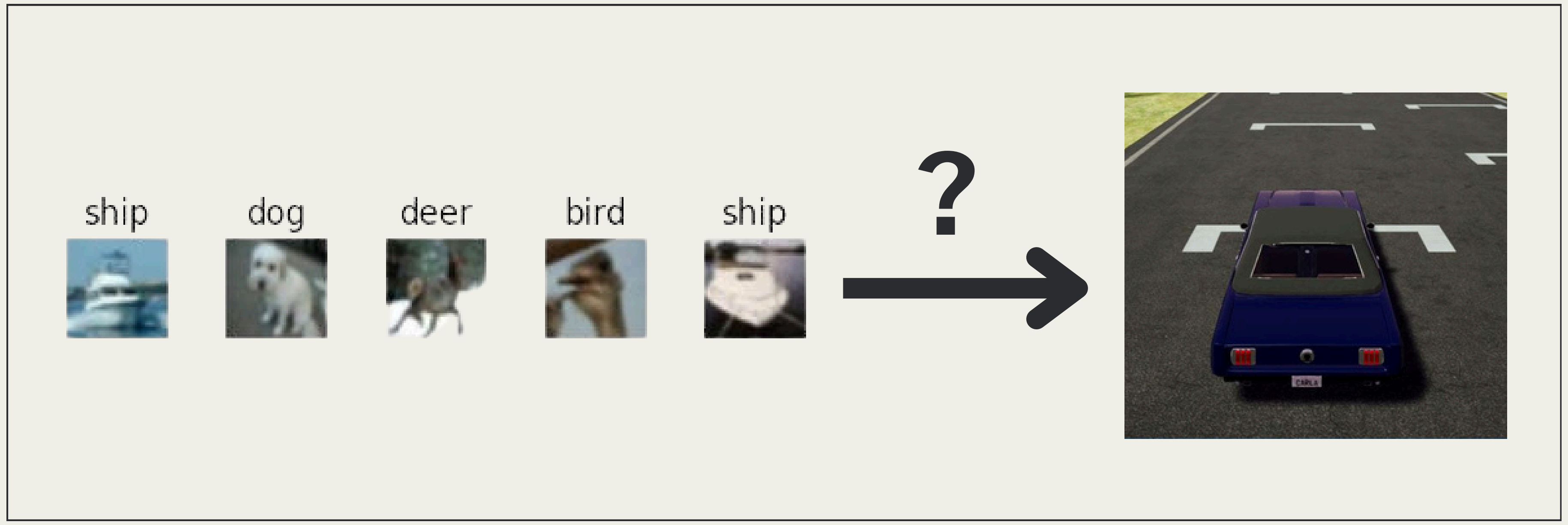
3. DOMAIN INDEPENDENCE

The big challenge: crossing between activity and perceptual domains.

Is it possible to have a feature representation that applies in a universal way?

- No such feature representation currently exists
- One possibility is to pick a universal representation (e.g., spectrogram) and transform all forms of data into it (Qiu et al. ICML 2021).

One is left minimizing domain dependence, rather than achieving strict domain independence

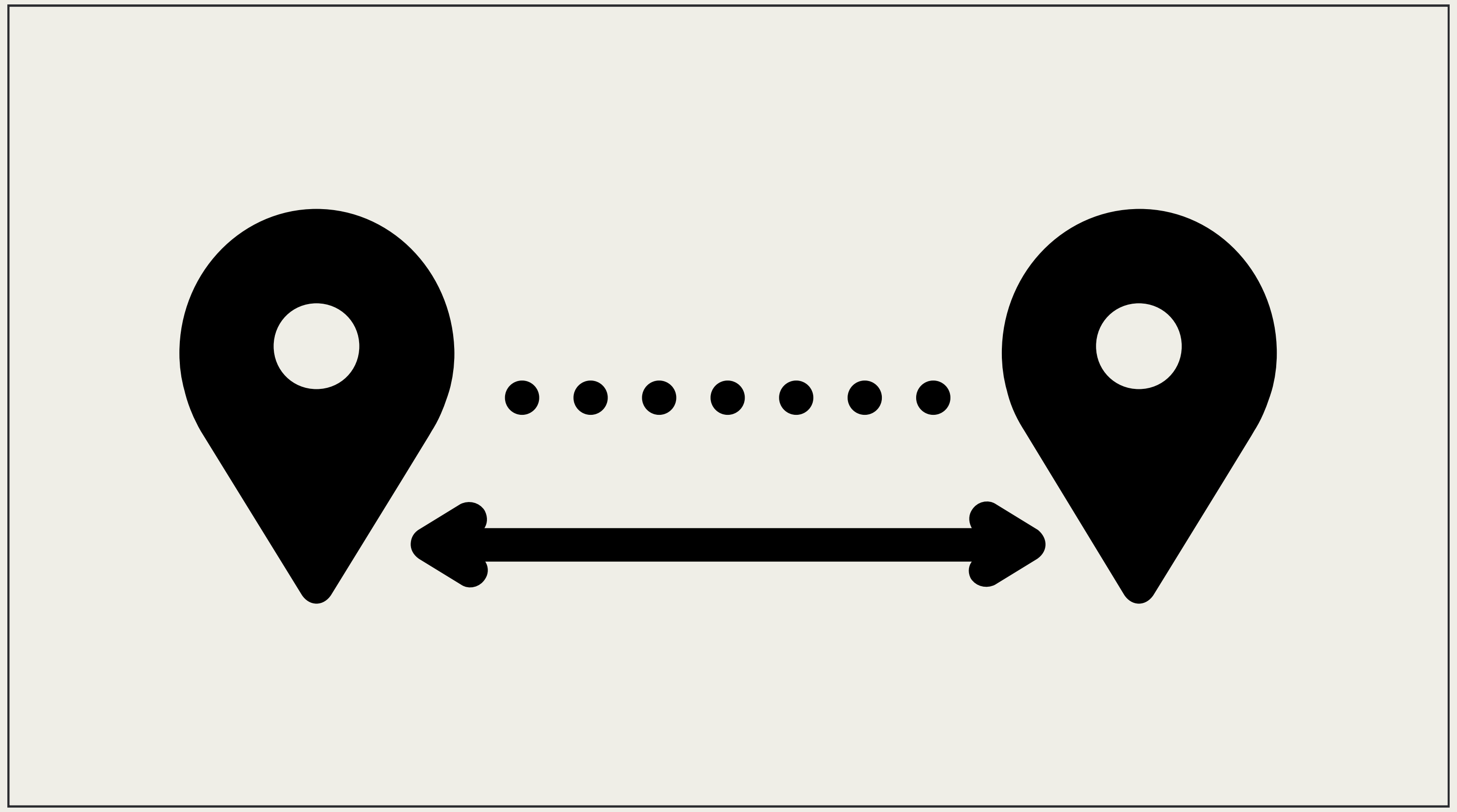


Qiu et al., "Neural transformation learning for deep anomaly detection beyond images," ICML 2021

4. BETTER REPRESENTATION FOR NOVELTY LEARNING

A few more thoughts on representation:

- There is an intrinsic link between the goodness of a representation and the ability to detect novelty.
- If known information can be clearly represented, then what is different from it can be discerned without significant effort.
- If there is too much aliasing between known and unknown information, false positives and false negatives will result.



Design of Agents

4. BETTER REPRESENTATION FOR NOVELTY LEARNING

Representation Edit Distance (RED) (Alspector AAAI Spring Symposium on Designing AI for Open Worlds 2022)

Measure of novelty that can be used by agents to adapt to it.

Change in information content in bit strings is measured by comparing pre- and post-novelty skill programs.

Bit string representations works across knowledge graphs, regressors, NNs, and even intuitive representations of knowledge.

Constraint: information theory setting; need good approximations to what should be formally optimal elements in the framework.

$$\boxed{RED^{\theta}_{a,T,C} = \frac{Cpretr\left(TrSol^{\theta}_{T} \middle| a_{t=0}\right)[]to]Cpost\left(Sol^{\theta}_{T}\right)}{Cpost\left(Sol^{\theta}_{T}\right)}}$$

Alspector, “Representation edit distance as a measure of novelty,” AAAI Spring Symposium on Designing AI for Open Worlds 2022

5. ROBUSTNESS TO NOVELTY VERSUS NOVELTY DETECTION AND CHARACTERIZATION

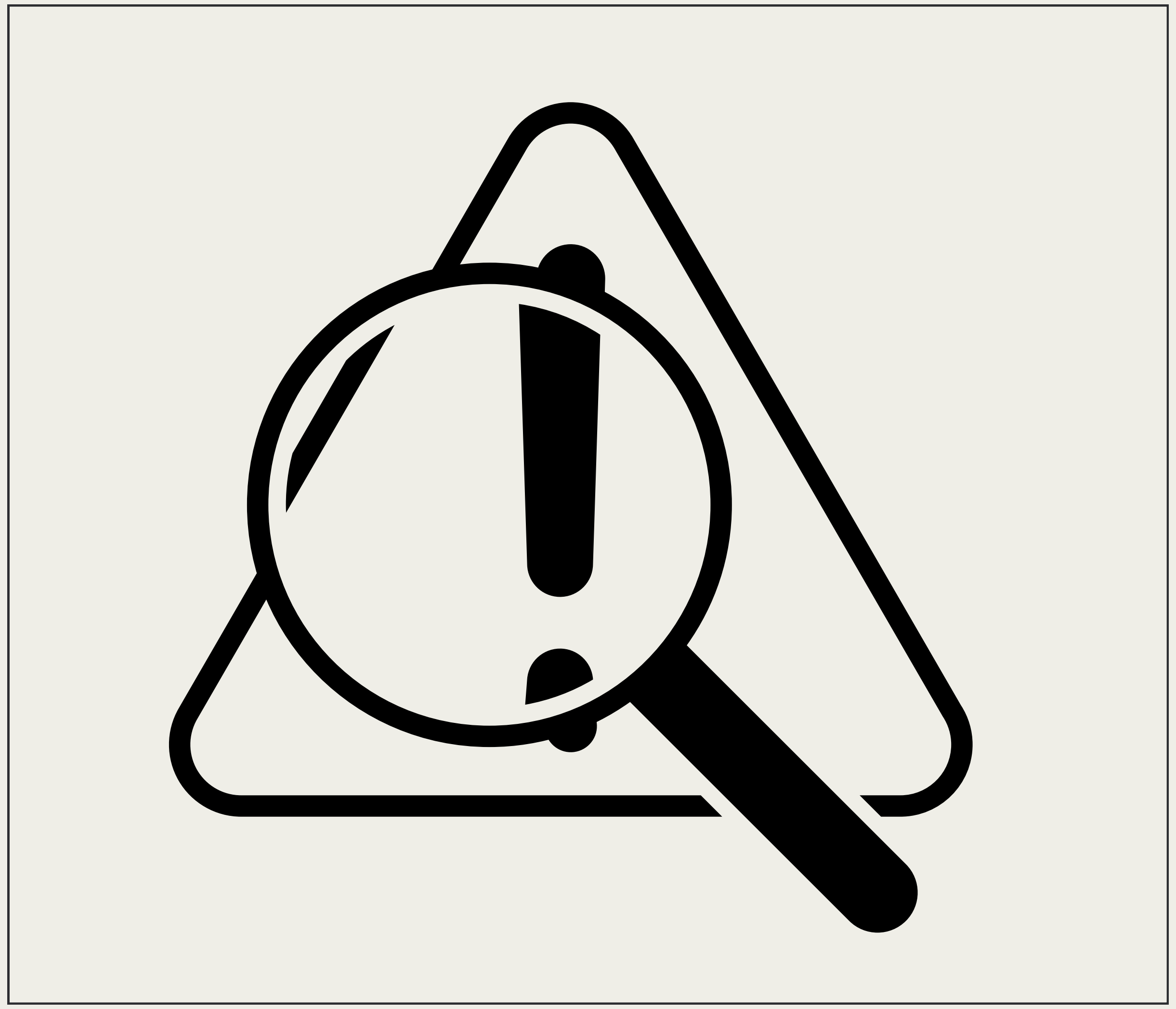
Huge amount of literature dedicated to novelty detection (hundreds of references as noted by Ruff et al. 2021)

- Not inherently useful by itself!

Novelty characterization is necessary to sort out nuisance novelty from novelties that must be processed by an agent

- Need a regret calculation for this

Novelty adaptation means an agent uses novel information to adjust its decision making process or as conditioning to ignore it.



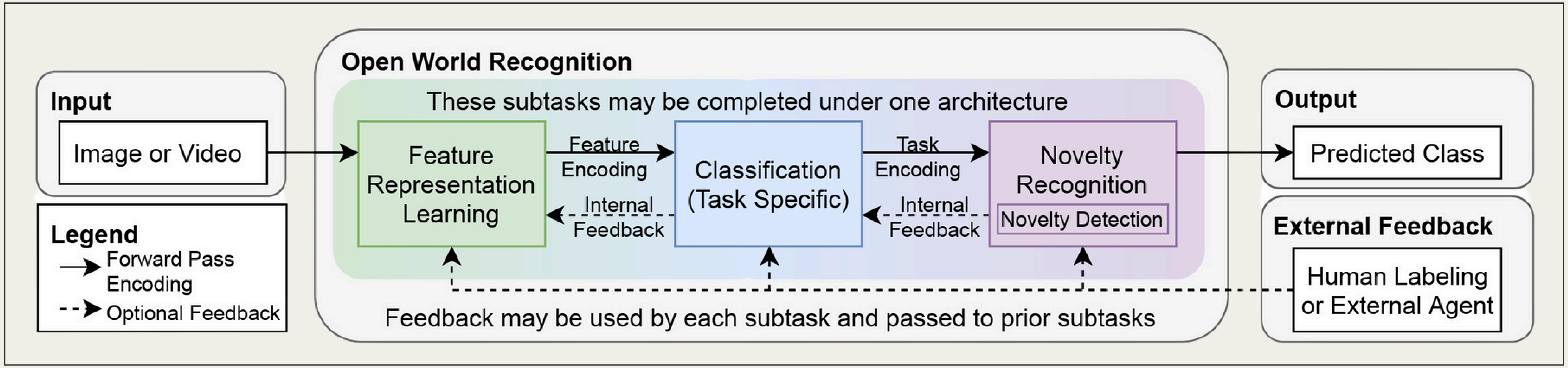
Design of Agents

Ruff et al., "A Unifying Review of Deep and Shallow Anomoly Detection," *Proc. of the IEEE* 2021

5. ROBUSTNESS TO NOVELTY VERSUS NOVELTY DETECTION AND CHARACTERIZATION

Strategy for perceptual domains: (1) feature extraction via NN, (2) novelty detection using extracted features, (3) clustering over novel features to identify new categories of novelty, (4) incremental learning to incorporate a new type of novelty back into the model.

Strategy for activity domains: (1) agent senses information from environment, (2) novelty detection using sensed information, (3) a detected novelty is characterized, (4) the agent's plan is revised to incorporate the novelty.

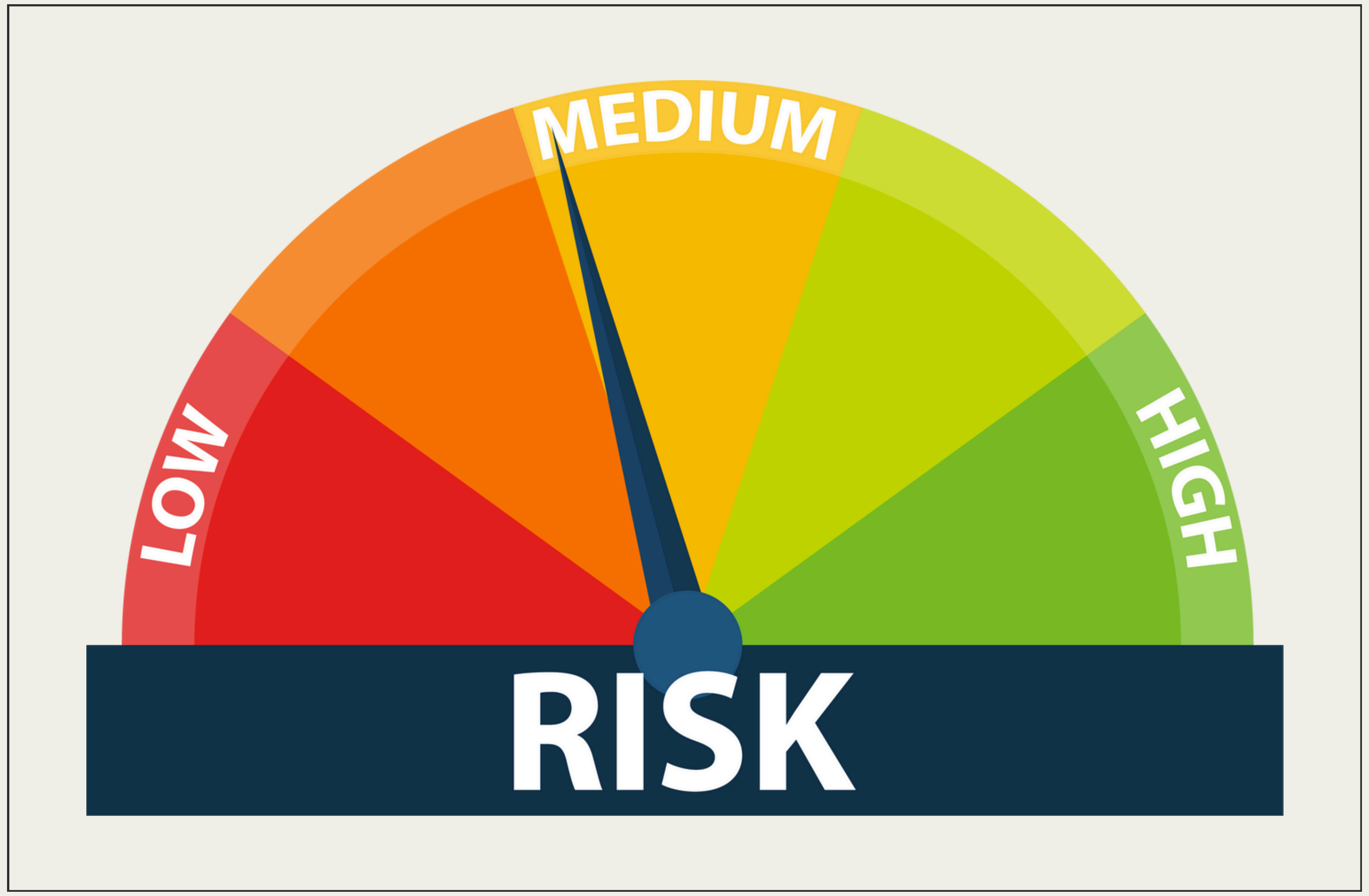


6. RISK-BASED REASONING

Novelty detection carries the **risk** of being **too insensitive** to hard-to-detect instances of novelty or being **overly sensitive** to all instances of novelty, leading to nuisance novelty.

Risk is abstracted by Boult et al. (AAAI 2021) as a regret operator on the world state, observed state or agent state.

- **Challenge:** choosing a threshold over the regret values



Design of Agents

Boult et al., “Towards a unifying framework for formal theories of novelty,“ AAAI 2021

7. SPECTRUM OF PARTIAL KNOWLEDGE THE SYSTEM DESIGNER HAS ABOUT NOVELTIES

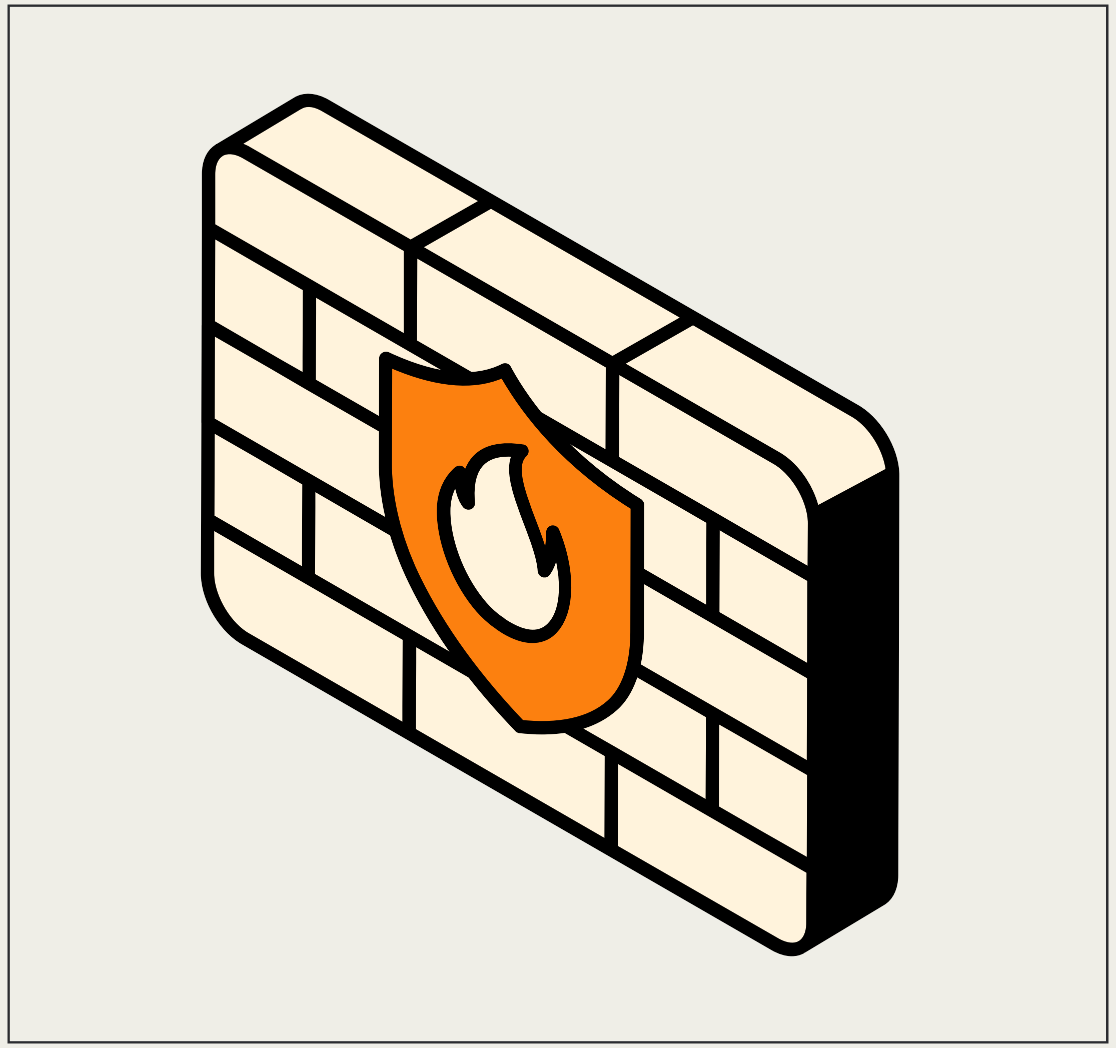
General problem in AI research: evaluations leak information

Novelty detection is particularly sensitive to information leaks between a novelty generator and novelty detector because novelty is unbounded.

- The search space is gigantic; leaks can artificially limit it.

A **firewall** should exist between generators and detectors.

Evaluation of Agents



7. SPECTRUM OF PARTIAL KNOWLEDGE THE SYSTEM DESIGNER HAS ABOUT NOVELTIES

Leak mitigation strategies:

- Have a sufficiently large validation set of data for the novelty detector to be trained with, with novelties that do not occur in the testing set for debugging purposes.
- Clearly describe everything considered to be known. This can be especially difficult with perception domains but is crucial since the problem becomes under-defined otherwise, making it impossible to tell the differences between nuisance novelty and managed novelty.
- During the creation of the novelty detector, ensure that the system is defined by looking for novelty rather than guessing a predefined set of potentially novel states.
- Ensure the problem space of potential novelties is large enough that it would be impossible to hand-code most of the novel states.

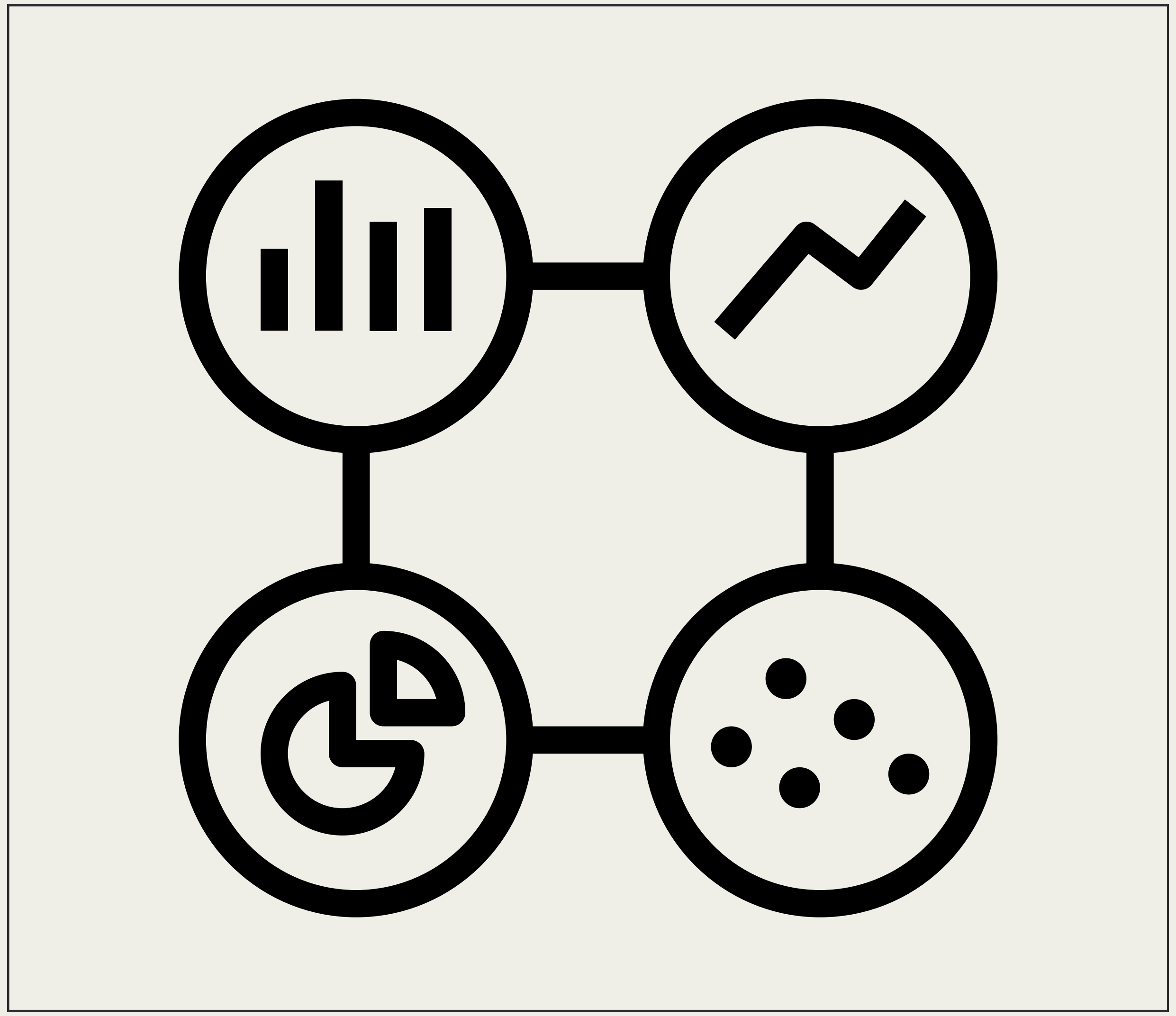
8. LACK OF MEASURES SPECIFIC TO OPEN WORLD LEARNING

Metrics that have proven inadequate for open world learning: Accuracy, Precision, Recall, F-1, AUC, and MCC

- All are limited to classification performance

Uncertainty induced by an encounter with novelty should be assessed

Domain-independent measure for estimating the **complexity** level of a domain provides a way to compare different domains (Doctor et al. AAAI Spring Symposium on Designing AI for Open Worlds 2022).



Evaluation of Agents

Doctor et al., “Toward defining domain complexity measure across domains,” AAAI Spring Symposium on Designing AI for Open Worlds 2022

# ACKNOWLEDGEMENTS

Notre Dame

Steve Cruz

Kitware

Chris Funk

Naval Research Laboratory

Katarina Doctor

Questions?