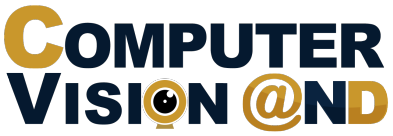
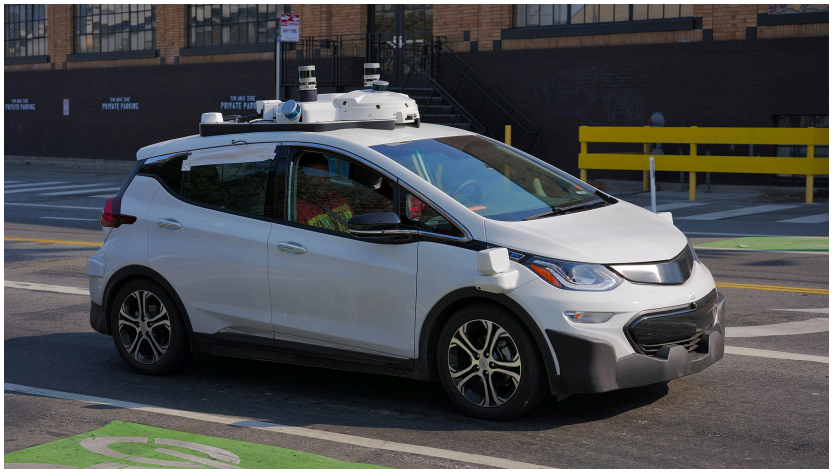
A Unifying Framework for Formal Theories of Novelty in Visual Perception

Walter J. Scheirer
University of Notre Dame



I was promised a fully self-driving car by 2018. What happened?



A Cruise Automation Chevrolet Bolt, third generation, seen in San Francisco. CC BY-SA 3.0 Dllu

- Novelty on the road still a significant impediment
- Even recording billions of miles of driving necessarily under samples the visual world

Novelty is a problem in any visual recognition domain once you move out of a dataset

Why is novelty a problem for visual recognition?

Open Space

Open Space

Walter J. Scheirer, Anderson Rocha, Archana Sapkota, Terrance E. Boult, "Towards Open Set Recognition," IEEE T-PAMI 2013

Many algorithms now addressing this

- P_I SVM: Jain et al. ECCV 2014
- W-SVM: Scheirer et al. IEEE T-PAMI 2014
- OpenMax: Bendale and Boult CVPR 2016
- Extreme Value Machine: Rudd et al. IEEE T-PAMI 2017
- CROSR: Yoshihashi et al. CVPR 2019
- Multi-task OSR: Oza and Patel arXiv 2019
- PROSER: Zhou et al. CVPR 2021

▪ ▪ ▪

Better, but still limited improvement in practice

What does it mean to be ‘novel’ though?

Is it a noticeable change in the world that does not impact an agent’s task performance?

How about a change that impacts performance but is not directly perceptible?

If the world has not changed but the agent senses a random error that produces an input that leads to an unexpected state, is that novel?

Towards a Unifying Framework for Formal Theories of Novelty

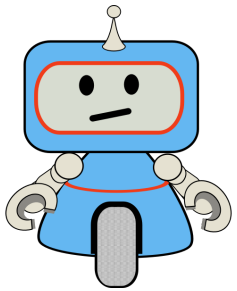
T. E. Boult , P. A. Grabowicz, D. S. Prijatelj, R. Stern, L. Holder, J. Alspector, M. Jafarzadeh, T. Ahmad, A. R. Dhamija, C. Li, S. Cruz, A. Shrivastava, C. Vondrick, W. J. Scheirer

AAAI 2021, https://ojs.aaai.org/index.php/AAAI/article/view/17766

A Unifying Framework for Novelty Theories

Problem: no consistent unified definition of novelty in lit.

Solution: formalize what it means for an input to be a novelty in the context of agents in artificial intelligence



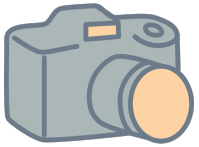
This means that many theories of novelty are possible, defined via a formal **framework**

Other theories subsumed by the framework

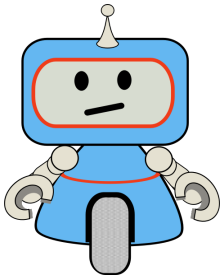
- Open Set Recognition: Scheirer et al. T-PAMI 2013
- Novelty Detection: M. Markou and S. Singh Signal Processing 2013
- Open World Learning: Bendale and Boult CVPR 2015
- Generative Theory of Novelty: Langley AAAI 2020

Three Spaces Where Novelty Occurs

The World Space $\mathcal{W}$

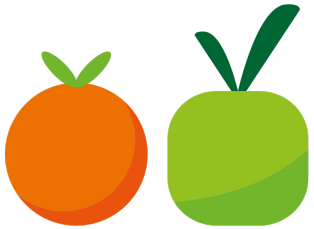


The Observation Space (i.e., What is Sensed) $\mathcal{O}$



The Agent (Perceptual or Activity) Space $\mathcal{A}$

A Couple More Definitions

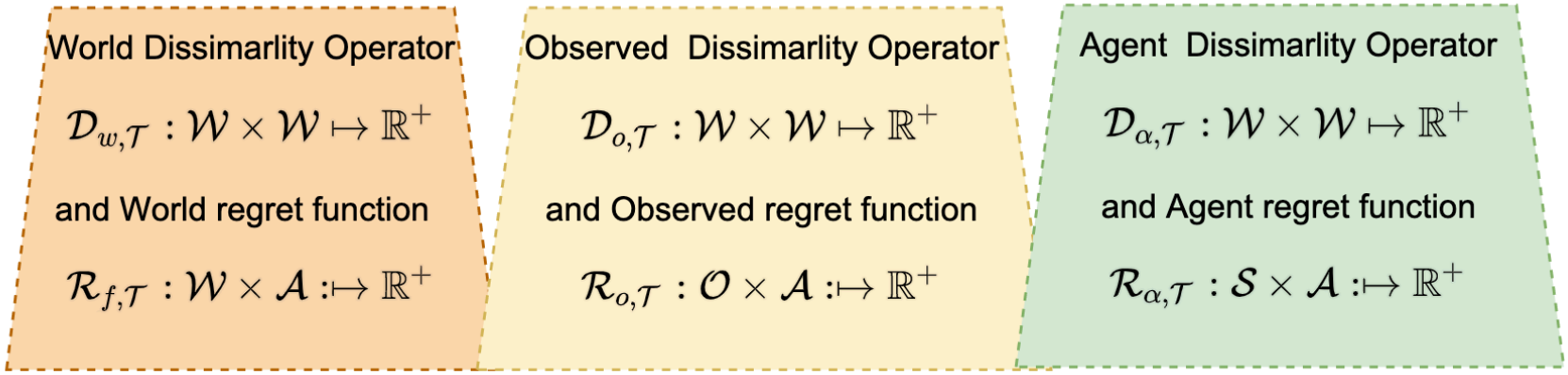


Dissimilarity between the potentially novel world state and the experience of some non-novel world states defines novelty

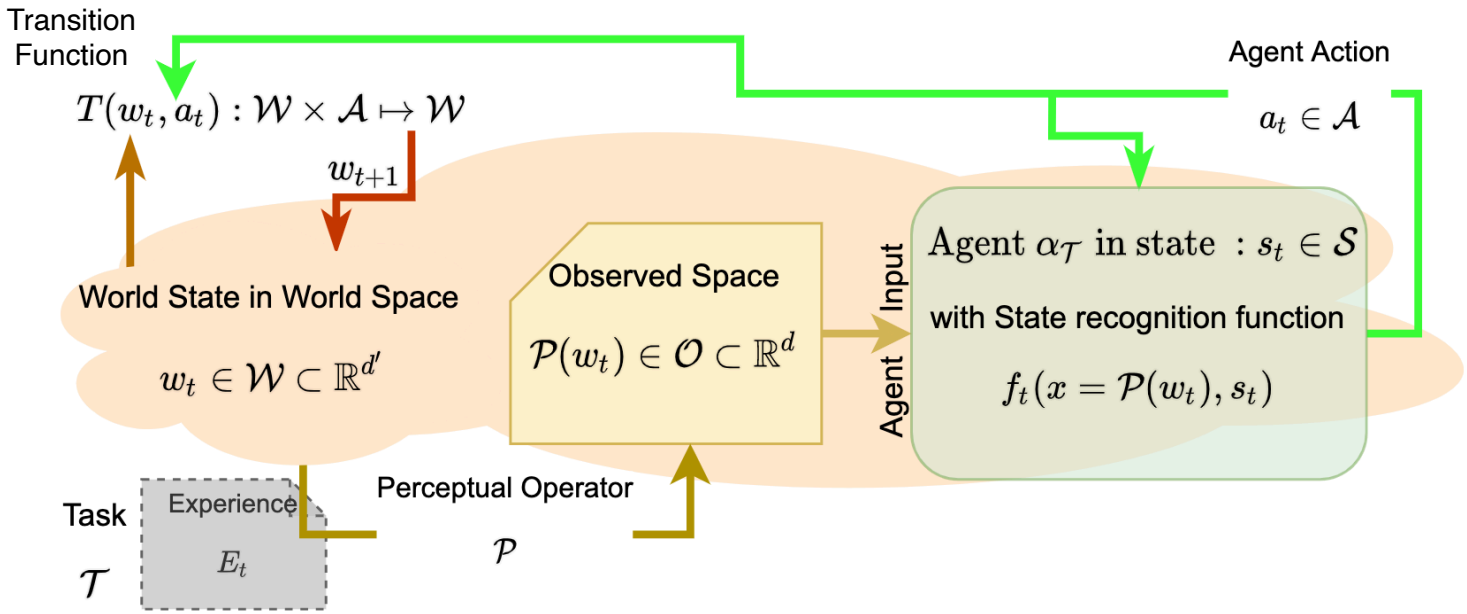
A **Regret** function assigns a penalty corresponding to a move to state $s_t \in S$

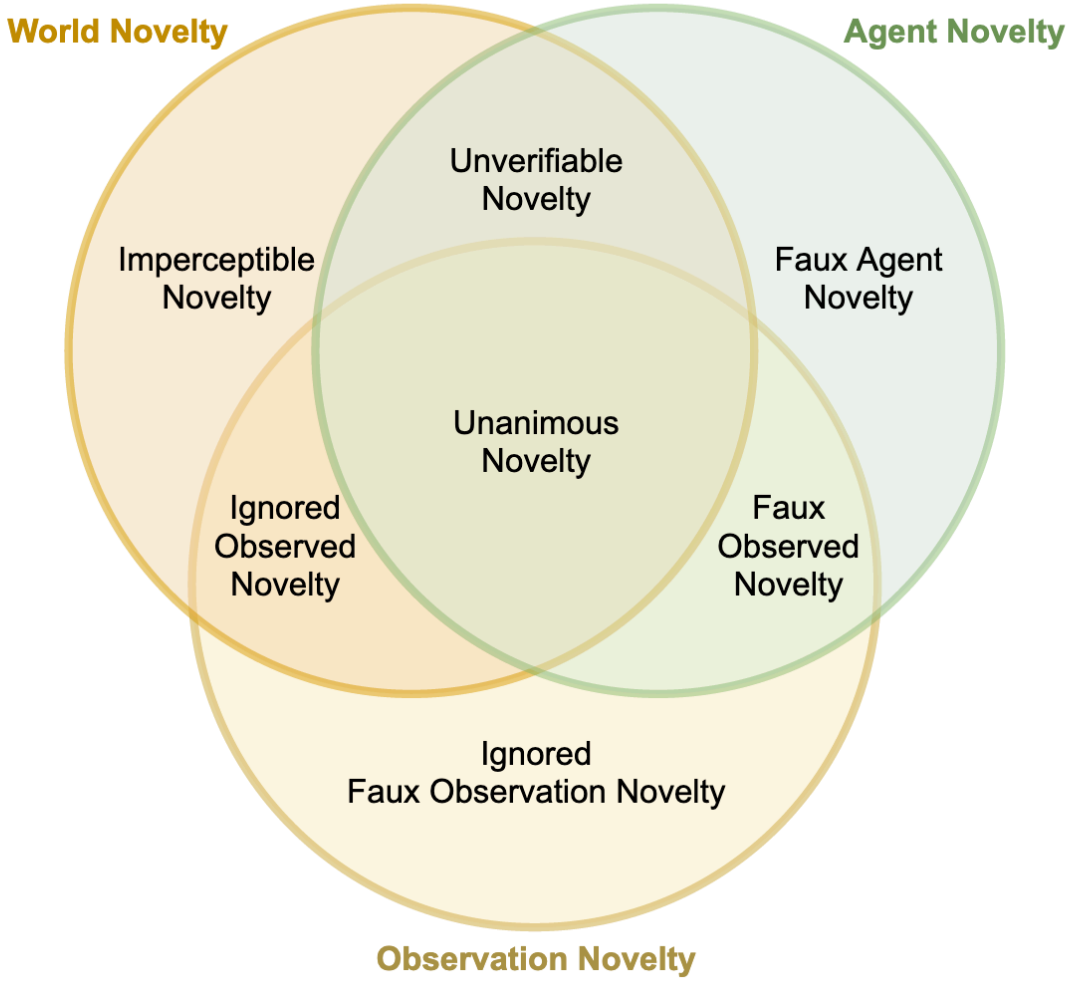
Dissimilarity and Regret Across Spaces

For a task τ:



Main Elements of an Implicit Theory of Novelty



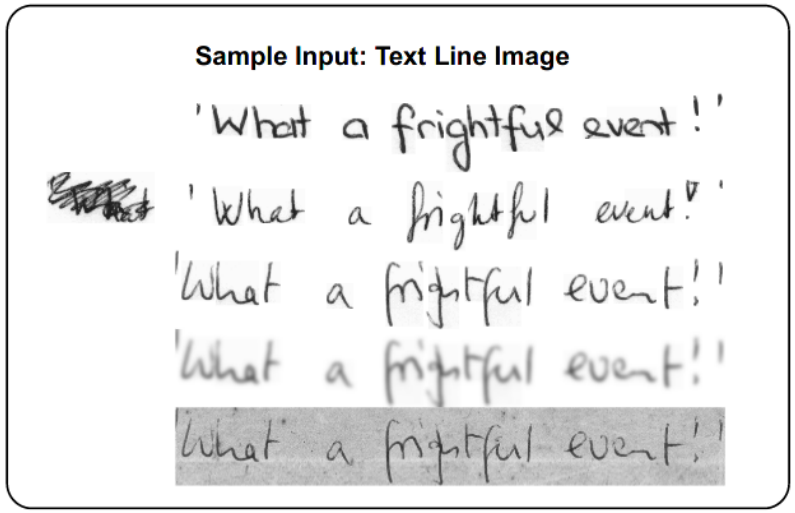

World Novelty
Agent Novelty
Unverifiable Novelty
Imperceptible Novelty
Faux Agent Novelty
Unanimous Novelty
Ignored Observed Novelty
Faux Observed Novelty
Ignored Faux Observation Novelty
Observation Novelty

Two Additional Novelty Subtypes

Managed novelty is a world novelty such that its implication on regret (performance) is minimal

Nuisance novelty is a novelty for which the world regret and the observation regret significantly disagree.

Example Task: Handwritten Document Transcription



Transcription	Writer ID	Appearance ID
'What a frightful event!'	031r	Standard Scan
# 'What a frightful event!'	031c	Standard Scan
'What a frightful event!'	031h	Standard Scan
'What a frightful event!'	031h	Blurred
'What a frightful event!'	031h	Antique

World spaces where novelty may occur:

- Transcription: Words, characters, glyphs
- Writer style
- Overall Document Appearance

Derek Prijatelj, Samuel Grieggs, Futoshi Yumoto, Eric Robertson, Walter J. Scheirer, "Handwriting Recognition with Novelty," ICDAR 2021.

Specification for Spaces

- Task τ is text transcription.
- A world $\mathcal{W}$ consists of a d'-dimensional space of pages of handwritten documents.
- An observation space O is a d-dimensional space that can encode all possible images of handwritten documents (i.e., the feature space).
- Perceptual operators P_t are optical sensors that capture a visible region of the world $\mathcal{W}_t$, where a time-step t results in a single image of a handwritten document.
- The perceptual operator may continue with feature extraction on the captured image to represent it as a feature vector of arbitrary dimensions to the agent in O.

Specification for Dissimilarity

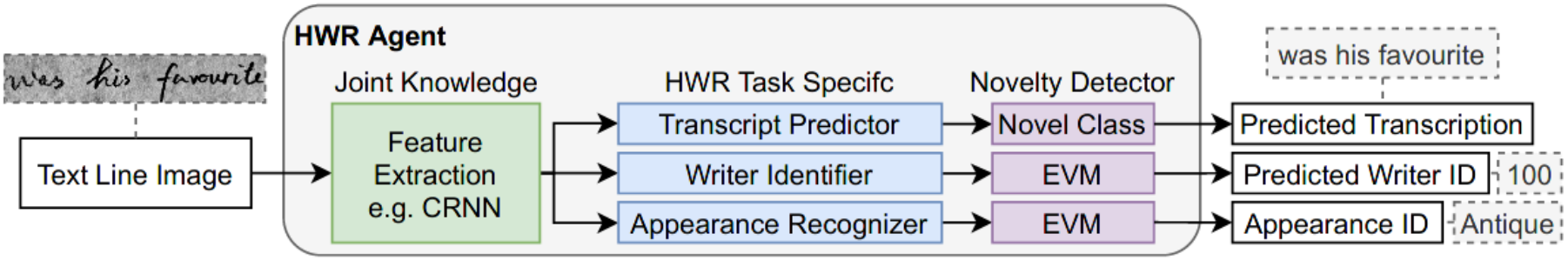
- The world dissimilarity functions $D_{w,\tau;E_t}$ and novelty threshold δ_w are determined by ground-truth labels.
  - The measurement of dissimilarity uses a distance measure (e.g., Euclidean) with a threshold determined by the probability distribution of the data.

- The dissimilarity functions $D_{o,\tau;E_t}$ and novelty threshold δ_o are determined by the agent's knowledge and design for the task.
  - In a learning-based agent, this is typically done via generalizing from the ground-truth in any available training or validation data.
  - These functions could make use of Euclidean distance or whichever distance measure suits the observational space.

Specification for Regret

- The world regret function $R_{w,\tau}$ is based on the error as measured in the world space given ground-truth labels.
  - For transcription, this could be Levenshtein Edit Distance, Character Error Rate, or Word Error Rate.
- The observational space regret function $R_{o,\tau}$ defines what the agent deems important to the task.
  - Embodied by the agent's internal model for the task, such as the loss function of a neural network or the likelihood calculation in a probabilistic model.

The Agent


was his favourite
Text Line Image
HWR Agent
Joint Knowledge
Feature Extraction e.g. CRNN
HWR Task Specifc
Transcript Predictor
Writer Identifier
Appearance Recognizer
Novelty Detector
Novel Class
EVM
EVM
was his favourite
Predicted Transcription
Predicted Writer ID
100
Appearance ID
Antique

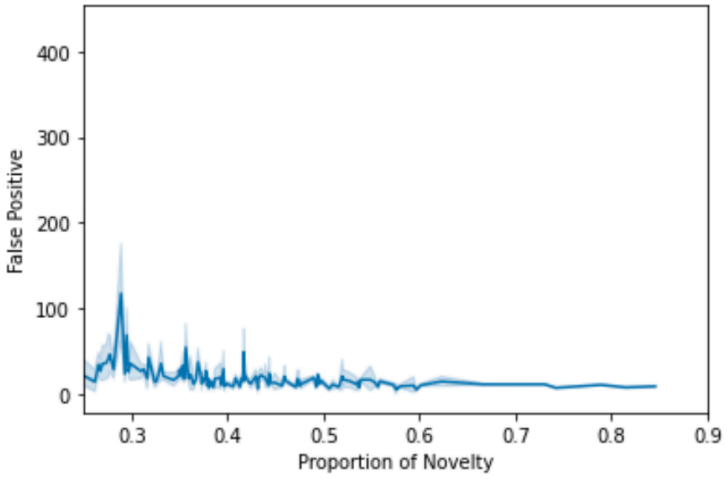
Predictions Made by the Theory

Dissimilarity implementation: 1 - cosine similarity

- E_t serves as the history of experience of the agent and plays a key role when a change in the world state is considered novel at a time step.
- The agent's training set and indirect information from the validation set are drawn from E_t, which informs the agent about how it should set its internal threshold δ_o.
- A world novelty may or may not effect the agent's performance on the task and the novelty's effect may vary in impact to task performance.

Predictions Made by the Theory

- Training data: IAM Offline Handwriting Dataset
- 55,000 novel writing samples used in evaluation were constructed from modified samples of IAM



What affects the false positive rate?

Factor	Sum of Squares	*df*	F	*p*
Distribution Type	1.058e+06	3	129.300	0.000
Level of Difficulty	7.456e+03	2	1.365	0.255
Location of Novelty	2.646e+06	1	969.301	0.000
Proportion of Novelty	5.848e+05	1	214.240	0.000
Residual	1.205e+00	44170		

ANOVA analysis of statistical influence given several test generating independent variables on false positive rate

Questions?