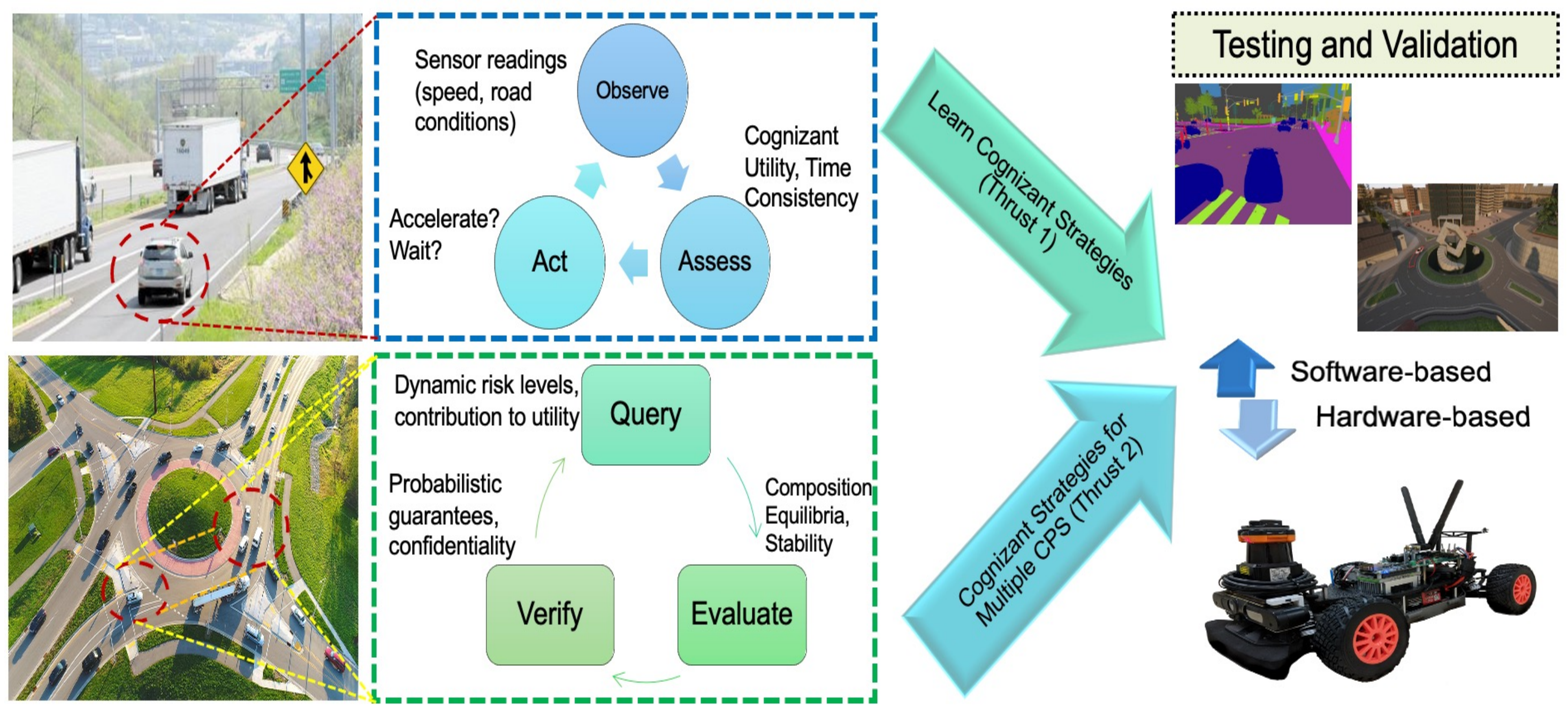
CRII: CPS: RUI: Cognizant Learning for Autonomous Cyber Physical Systems

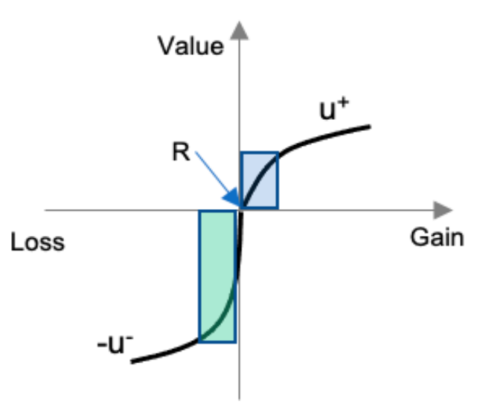
PI: **Bhaskar Ramasubramanian**, Assistant Professor, Western Washington University

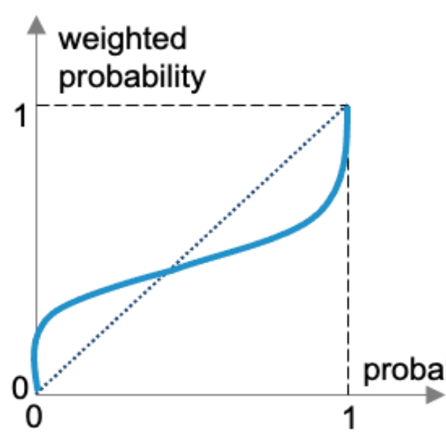
Project URL: https://sites.google.com/view/safeaicpslab/research

Objective and Setup:

- Develop cognizant learning framework for CPS grounded on autonomous driving
- Dynamic environments with multiple decision makers
- Risk-sensitive behaviors


Sensor readings (speed, road conditions)
Observe
Cognizant Utility, Time Consistency
Accelerate? Wait?
Act
Assess
Dynamic risk levels, contribution to utility
Query
Probabilistic guarantees, confidentiality
Composition Equilibria, Stability
Verify
Evaluate
Learn Cognizant Strategies (Thrust 1)
Cognizant Strategies for Multiple CPS (Thrust 2)
Testing and Validation
Software-based
Hardware-based





Challenges:

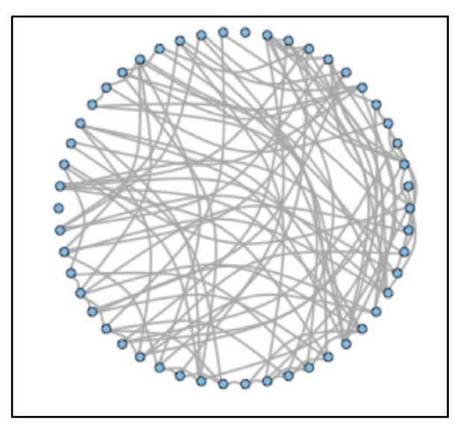
- Decision making when CPS and human share environment
- Model human behaviors
  - Deviation from reference, loss aversion
  - Improper quantification of probabilities
  - Maintain privacy of decision making
  - Interactions among multiple agents
- Maximizing average utility not adequate

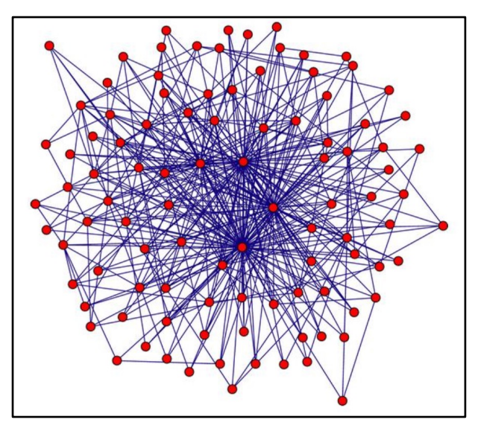
Scientific Impact:

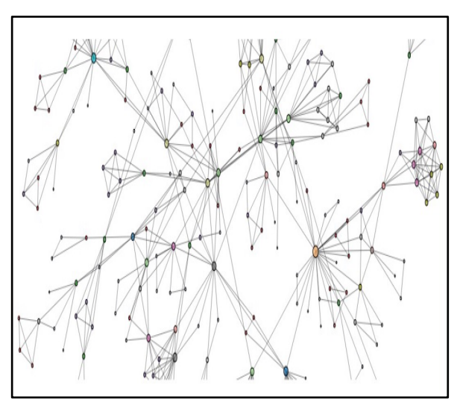
- Improve CPS-human interactions by unifying perspectives from learning, control, and behavioral economics
- New paradigm to learn behaviors consistent with human preferences using prospect theory (PT) [1, 2]
- Multi-modal defenses against vulnerabilities in GenAI [3]

Solution:

- Multi-agent PT-based RL algorithms [1]
- Cognitive bias-aware solutions in opinion dynamics models [2]
- Effects of prompting strategies in LLMs and backdoors in DNNs [3]
- Extensive experimental evaluations







Initial Predisposition	WS: % of Final Opinions > 0.5		BA: % of Final Opinions > 0.5		FB: % of Final Opinions > 0.5	
	Proposed model	Baseline model	Proposed model	Baseline model	Proposed model	Baseline model
$Unif(-1,0)$	**18.5%**	0	**12.1%**	0	**7.91%**	0
$Unif(0,1)$	**21.6%**	0.51%	**22.1%**	0.65%	**15.26%**	0
$Unif(-1,0.5)$	**14.5%**	0	**14.7%**	0.1%	**9.49%**	0
$Unif(-1,0.5)$ Or $Unif(0.5,1)$	**17.7%**	6.5%	**17.9%**	5.6%	**12.13%**	0

Broader Impacts: Research

- Improved CPS-human interactions grounded on large networked systems
- Reasoning about risk-sensitive decision-making
- Students at public PUI exposed to research
- Vertically integrated team
- One student awarded **WWU Summer Research Award**

Broader Impacts: Education

- **UG students** engaged in algo design, experiments [1]
- Eight UG students supported, including one woman
- **New UG courses** on AI-RL and CPS included modules on risk-aware learning and CPS
- UG students present posters at WWU Scholars Week
- Jobs at Boeing, PACCAR, TI

Dissemination:

1. **D. Danis, P. Parmacek, D. Dunajsky**, **B.R.**, *Multi-agent RL with cumulative prospect theory*, SIAM Conf. on Control (SIAM CT), 2023.
2. A. A. Maruf, L. Niu, **B.R.**, A. Clark, R. Poovendran, Learning Dissemination Strategies for External Sources in Opinion Dynamic Models, IJCAI 2023.
3. **B.R.**, et al., BadChain (ICLR 24); FedGame (NeurIPS 23); MDTD (CCS 23); LDL (AsiaCCS 23)

Electrical and Computer Engineering, Western Washington University

Contact: ramasub@wwu.edu

Award ID#: CNS 2153136