A Systems Motivation for the Design of Human-AI Interdependence

Tyler Cody, Stephen Adams, Peter Beling

Systems and Information Engineering, University of Virginia

AAAI 2020 Spring Symposium

March 24, 2020

Executive Summary

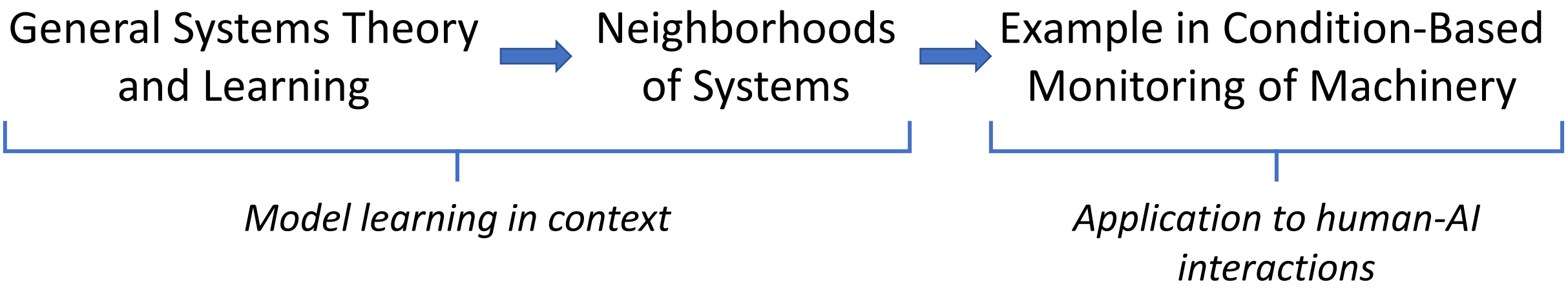
Assertion

We can apply general systems theory to model learning in context and elicit pertinent human-AI interdependencies.

Approach

We develop a **general goal for learning systems** using general systems theory, and show how **designing for that goal creates an awareness and knowledge about human-AI interdependence** in condition-based maintenance systems.

Outline


General Systems Theory and Learning
Neighborhoods of Systems
Example in Condition-Based Monitoring of Machinery
Model learning in context
Application to human-AI interactions

Systems Approach Has a Focusing Effect

Traditionally studies are

…human-centric looking outward from the cognitive and affective nature of humans, or…

- Lubars, B. and Tan, C. *"Ask not what AI can do, but what AI should do: Towards a Framework of Task Delegability."* NeurIPS 2019.

…AI-centric looking outward from the algorithmic nature of AI.

- Carroll, M. et. al. *"On the utility of learning about humans for human-AI coordination."* NeurIPS 2019.

Systems-centric approaches use a top-down, holistic perspective from which context can be taken into consideration.

Systems Theory

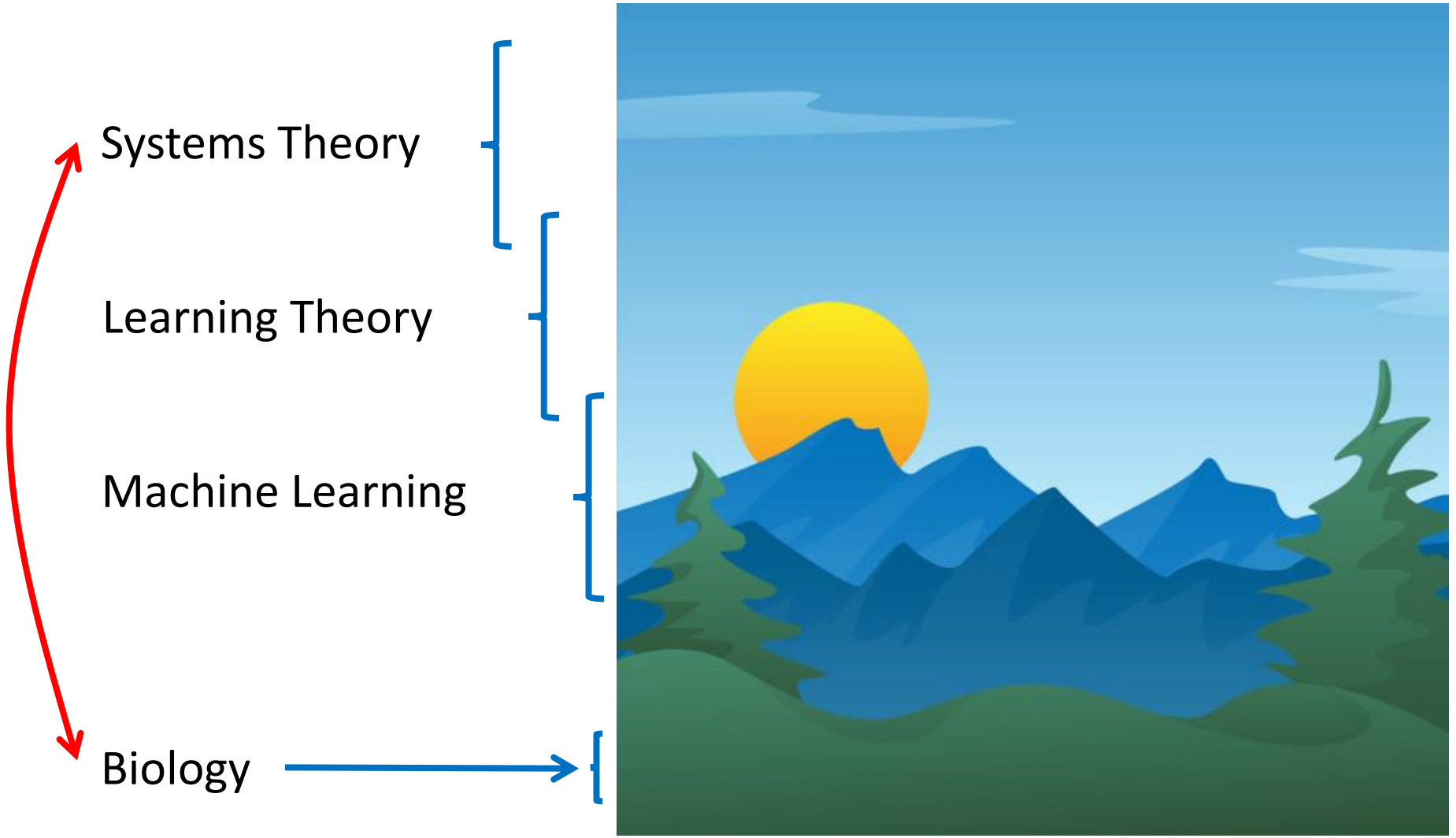
1. Background
2. Mathematics

What is systems theory?

"… there exists models, principals, and laws that apply to generalized systems or their subclasses, irrespective of their particular kind, the nature of their component elements, and the relationships of "forces" between them. It seems legitimate to ask for a theory, not of systems of a more or less special kind, but of universal principals applying to systems in general."

- Ludwig von Bertalanffy, General System Theory (1968)

Levels of Abstraction



Important Voices in Mathematical Systems Theory

Cybernetics

Ashby, *Principles of the Self-Organizing Dynamic System* 1947

Weiner, *Cybernetics: Or Communication and Control…* 1948

Studied relations between living and mechanical systems to draw conclusions about regulation of systems

Founder of General Systems Theory

Descriptive General Systems Theory

Bertalanffy, *General Systems Theory* 1949

Bertalanffy et al., *Society for General Systems Research* 1954

"The formal correspondence of general principles, irrespective of the kind of relations or forces between the components, leads to … a new scientific doctrine…"

Mathematical General Systems Theory

Forrester, *Industrial Dynamics* 1961 → Mathematical study of system dynamics

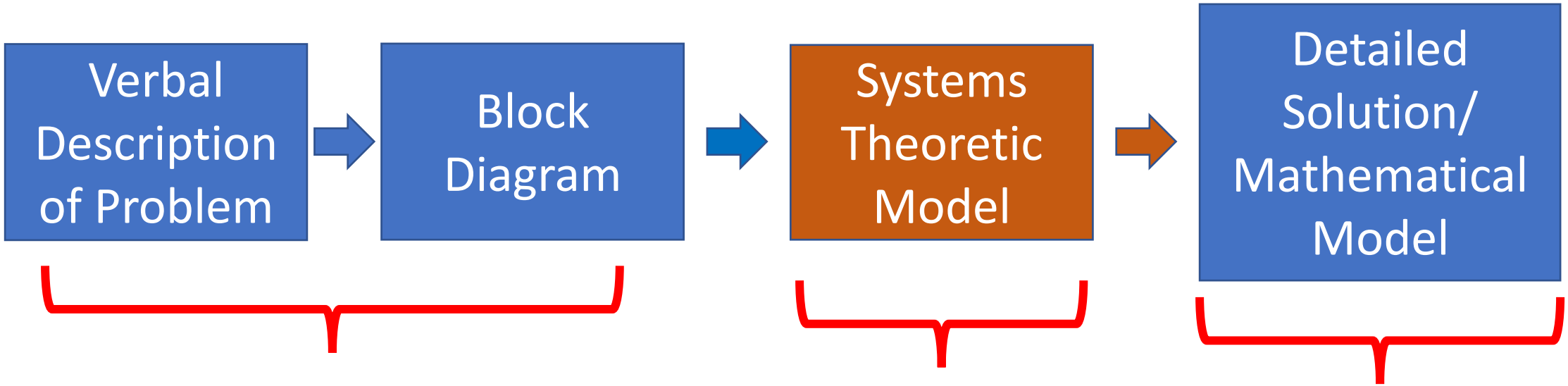
Mesarovic, *Views on General Systems Theory* 1964

Wymore, *A Mathematical Theory of Systems Engineering* 1968

Set theoretic interpretations of systems

Klir, *An Approach to General Systems Theory* 1968 → Systems as collection of traits

Between Block Diagrams and Detailed Models



What is the learning problem, and where does it reside within the context of our system?

How do we represent it in explicit, mathematical terms?

How do we algorithmically solve the problem?

Systems Theory

1. Background
2. Mathematics **– Formulating a General Systems Goal**

Mesarovician Systems Theory (MST)

- **Definition**. *System*.

A general system is a relation on non-empty (abstract) sets,

$$S \subset \times \{V_i : i \in I\}$$

where $\times$ is the Cartesian product, I is the index set, and V_i are the component sets.

- **Definition**. *Input-Output System*.

An input-output system is a general system where the component sets can be partitioned into an input object and output object,

$$X = \times \{V_i : i \in I_x\} \text{ and } Y = \times \{V_i : i \in I_y\}$$

where $I_x \cup I_y = I$. Thus,

$$S \subset X \times Y$$

MST + Empirical Risk Minimization

- **Definition**. *Learning System*.

A learning system is an input-output system,

$$S: X \to Y$$

with a sample D and a learning algorithm $A: D \to f^{\theta}$, where $f^{\theta}: X \to Y$ is a parameterized mapping.

- **Definition**. *An Empirical Risk Minimization Learning System*.

An empirical risk minimization learning system is a learning system where D is an i.i.d. sample of l input-output observations,

$$A: D \to f^{\min_{\theta} R_{emp}(\theta)}, \text{ where, } R_{emp}(\theta) = \frac{1}{l}\sum_{i=1}^{l} L(y_i, f^{\theta}(x_i))$$

and L is a loss function.

Neighborhoods, Random Processes, and Life Cycle

A learning algorithm A is a map,

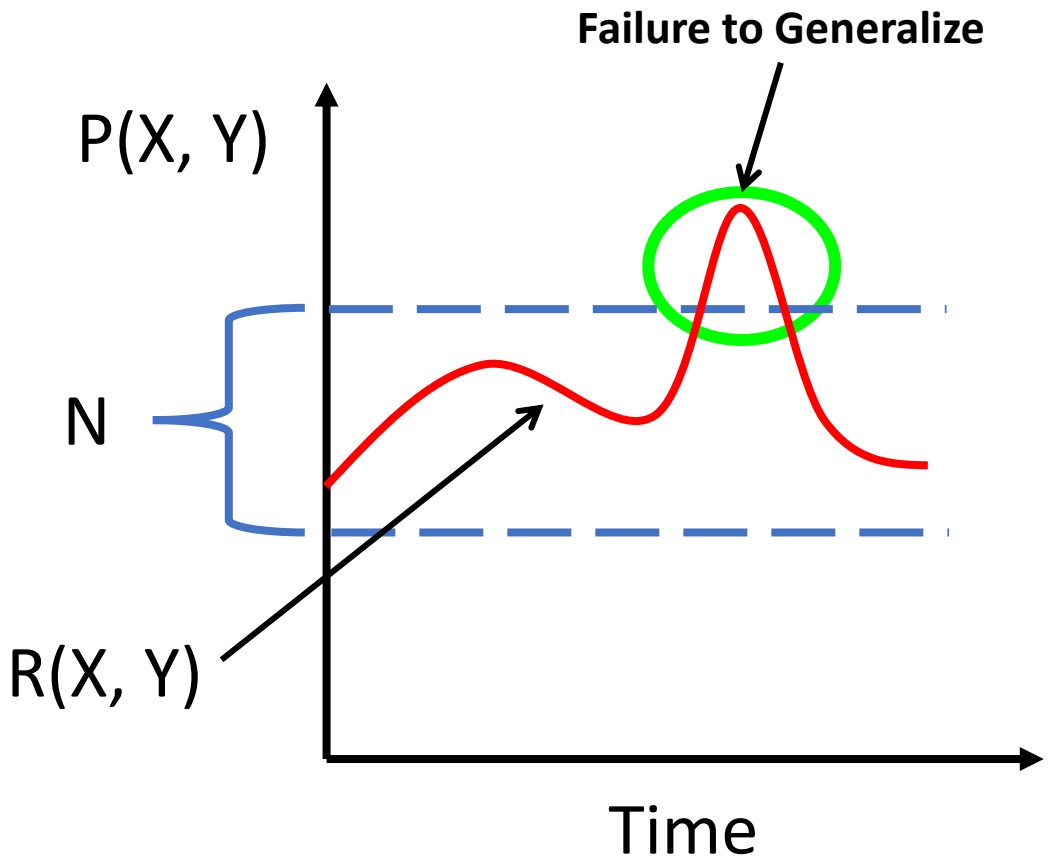
$$A: D \rightarrow f^{\theta}\ , f^{\theta}: X \rightarrow Y.$$

Given an evaluation function, $v: f^{\theta} \rightarrow \mathbb{R}$, and a real threshold ϵ, we are interested in identifying the neighborhood,

$$N = \{P(X,Y) | v(f^{\theta}) \geq \epsilon\}$$

System behavior is captured by the random process,

$$R(X,Y) = \{P_t(X,Y) | t = 1, \ldots, T\}$$

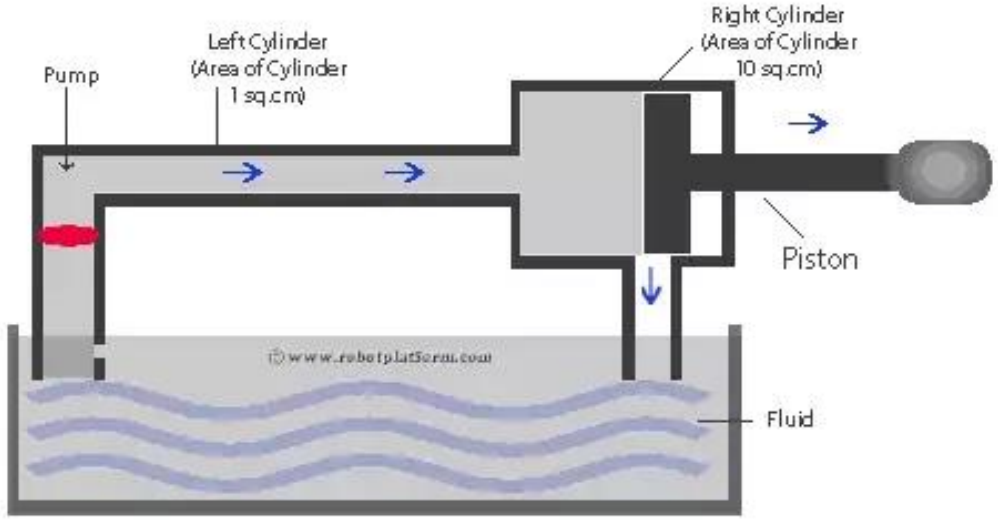


Human-AI Interdependence in Condition-Based Maintenance

1. Background
2. Case

Hydraulic Actuators

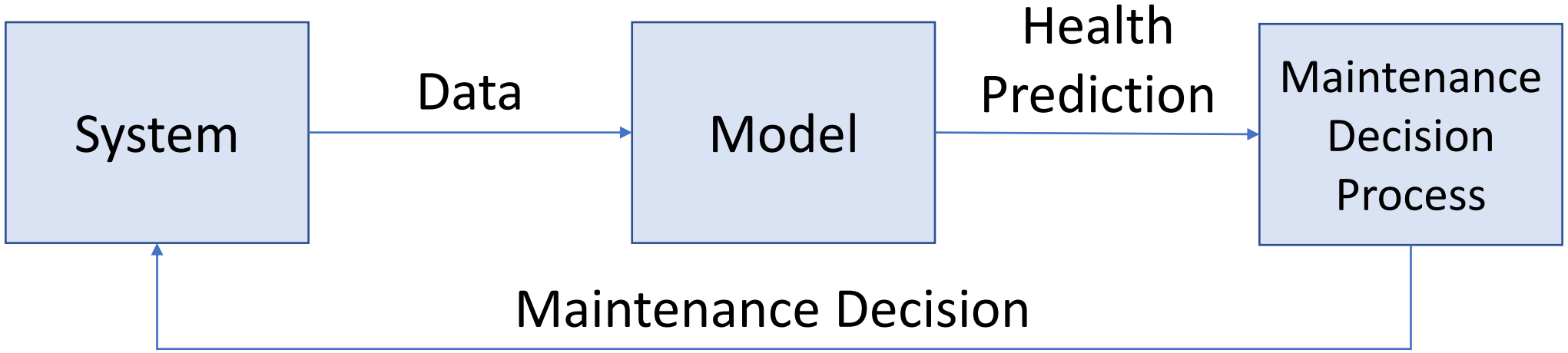
- Use hydraulic power to facilitate mechanical operation
- In the Navy, several hundred can be used on a vessel for critical and non-critical functions
- Currently, most are maintained on a time based schedule, but there is a desire to move towards condition based maintenance



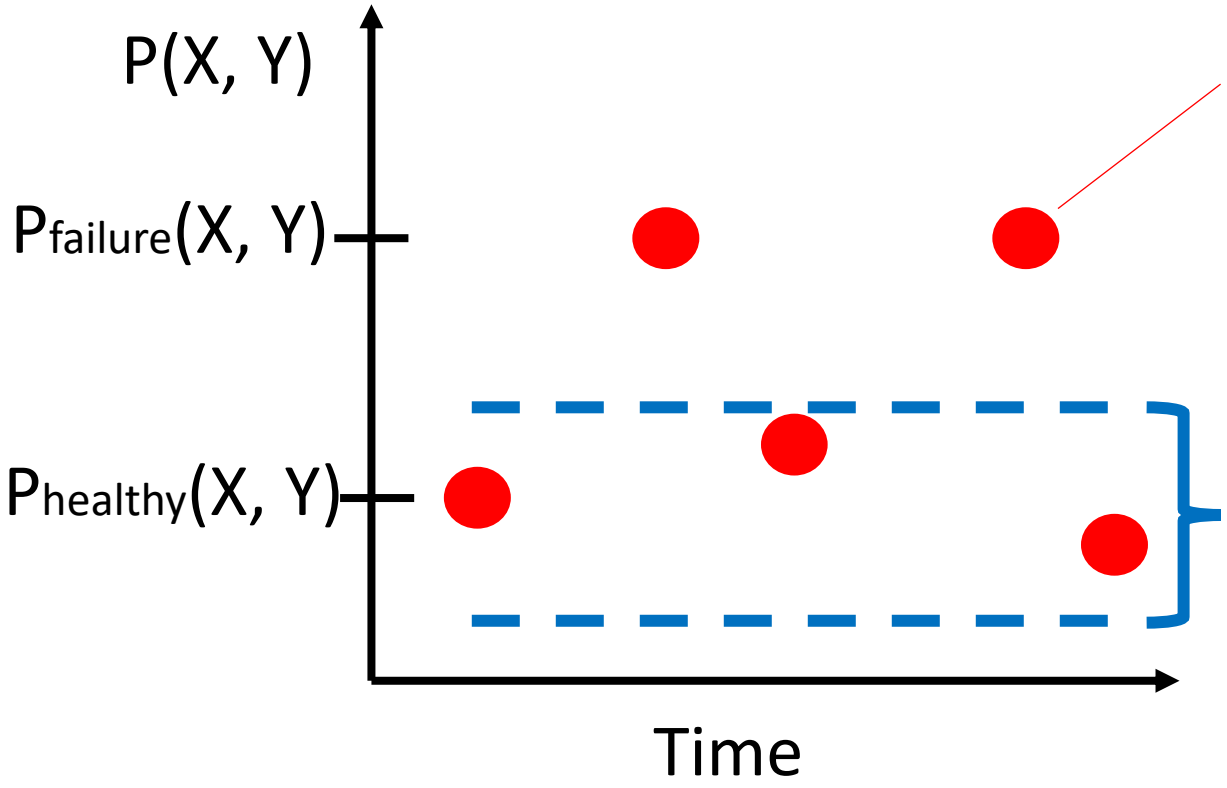
Hydraulic Actuator

Condition Based Maintenance

- Idea
  - Use knowledge/predictions about current and future health of a system to plan its maintenance
- Data-driven approaches use models to evaluate sensor readings



Goals for and Design of Rebuild Procedures



How do the technicians actions influence the neighborhood of behaviors we see over a system's life cycle?

Conclusion

- Systems theory can model learning in its systems context
- General goals can direct study of human-AI interdependence
- Considering system goals and system design has a focusing affect on the study of human-AI interdependence, directing studies towards operational awareness and understandings of interdependence

Questions?

tmc4dk@virginia.edu