

**CS 295B/CS 395B**  
**Systems for Knowledge**  
**Discovery**

Rethinking Industrial Roles



The University of Vermont

---

---

## End-of-Semester Logistics: Schedule

- **Today:** Last traditional lecture
- **Monday:** Last paper presentations/reviews (reviews due Sun. night!)
- **Wednesday:** Tentative **Online Only** guest lecture/Q&A

---

---

**Alex Passos**, former tech lead TensorFlow

scikitlearn OG contributor/author

Expertise: ML systems at scale



Interview about TF eager:

<https://www.youtube.com/watch?v=Ow3GkOUAfuM>



---

---

# End-of-Semester Logistics: Schedule

- **Today:** Last traditional lecture
- **Monday:** Last paper presentations/reviews (reviews due Sun. night!)
- **Wednesday:** Tentative **Online Only** guest lecture/Q&A
- **Friday:** Final class
  - Future work in this area (~30mins)
  - Course evaluations (20mins, I will leave the room)

***Bring your laptop if you have not yet completed them!***

# End-of-Semester Logistics: Grad Final Projects

**Reminder:** Upload to BB by 10:30am Thurs Dec 16

Guidelines to be released this weekend

Email me or self-schedule to meet and discuss:

- Teams -> Calendar icon (left panel)
  - Click on (soonest) time you want
  - Scheduling Assistant tab

**Reminder: I WFH on Tues & Thurs**

The screenshot displays the Microsoft Teams Scheduling Assistant interface. The 'Scheduling Assistant' tab is highlighted with a red circle. The interface shows a calendar view for Monday, December 6, 2021, with a meeting slot highlighted from 3:00 PM to 3:30 PM. The meeting details include the title 'New meeting', time zone '(UTC-05:00) Eastern Time (US & Canada)', and suggested times. The interface also shows a list of attendees, including Emma Tosch and Lisa Dion, and options to add required or optional attendees and locations.

---

---

## Topics today

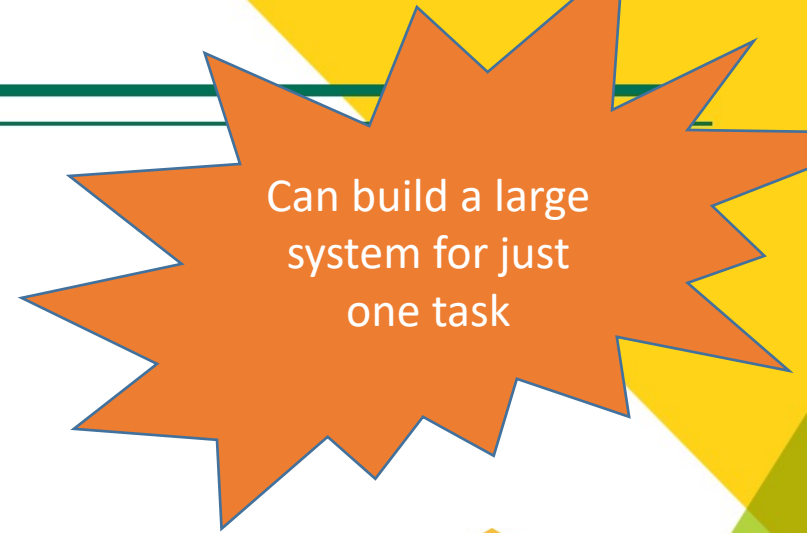
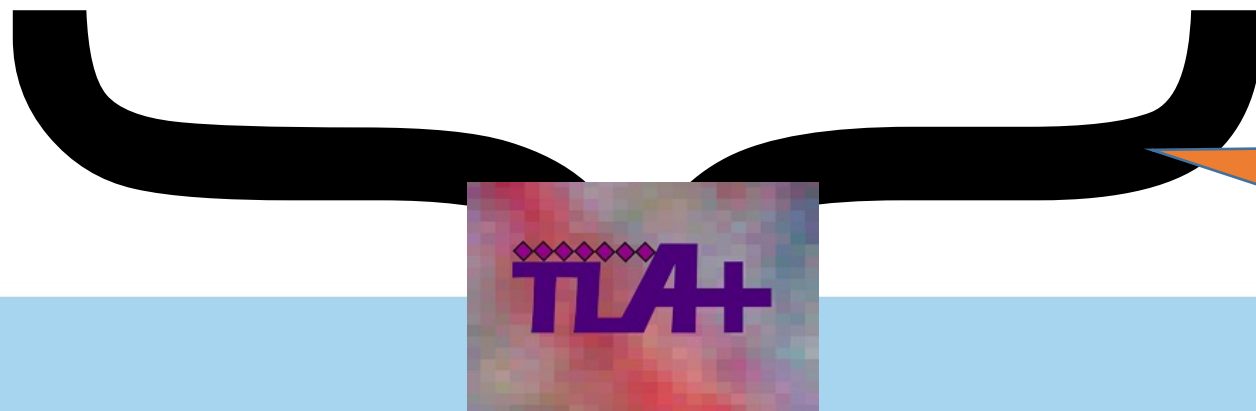
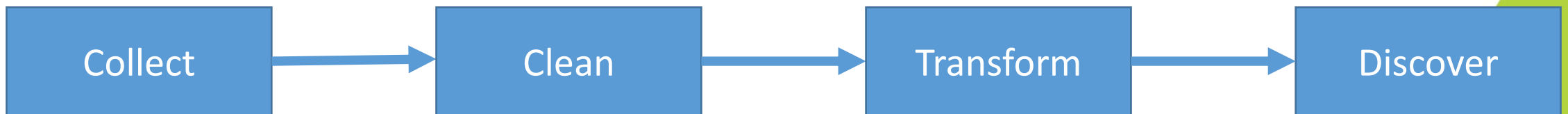
- 1) Current role of modern industrial systems research
- 2) Building tools vs. Governing usage of those tools  
*i.e., from systems to knowledge discovery*

# Current Role of Modern Industrial Systems Research

# Last Class

Most systems-y papers we have seen so far

*Where's the KDD?*





---

---

# What industry adds

**Scale** – simply more users/machines

**Applications/Applicability** – ability to evaluate over more data/scenarios

**New problems** – scale + real-life use can highlight problems not seen before

**Engineering clout** – i.e., why these systems are actually used

**Money** – both internally (to fund this work) and externally (academic grants)

---

---

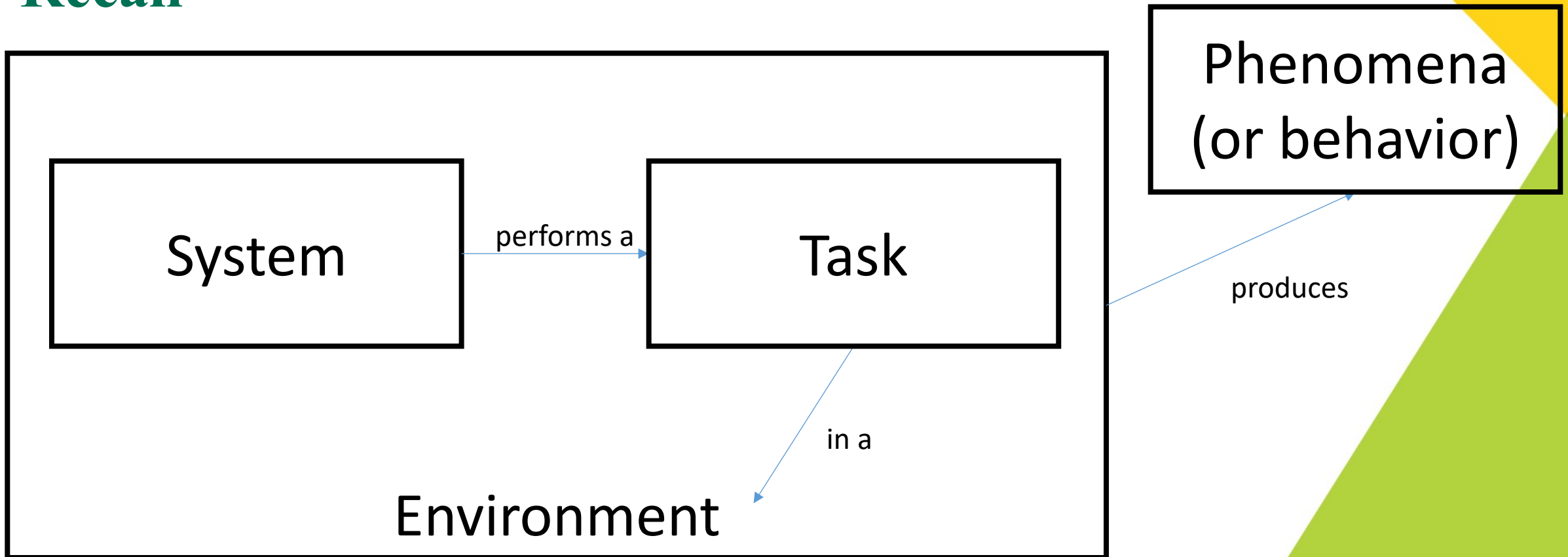
# Industrial complications

**Money** – a blessing and a curse

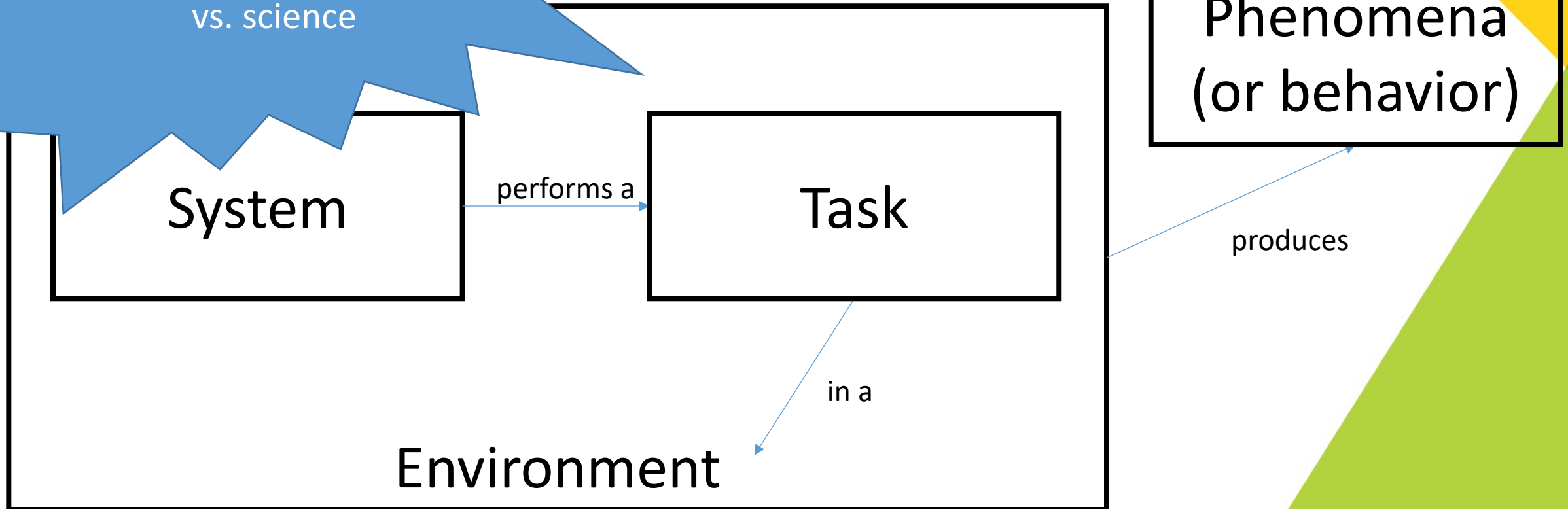
- Resource that powers all research (unless you are independently wealthy)
- Can drive interest away from basic research toward applied

**Boards of directors, public relations** can influence study of phenomena

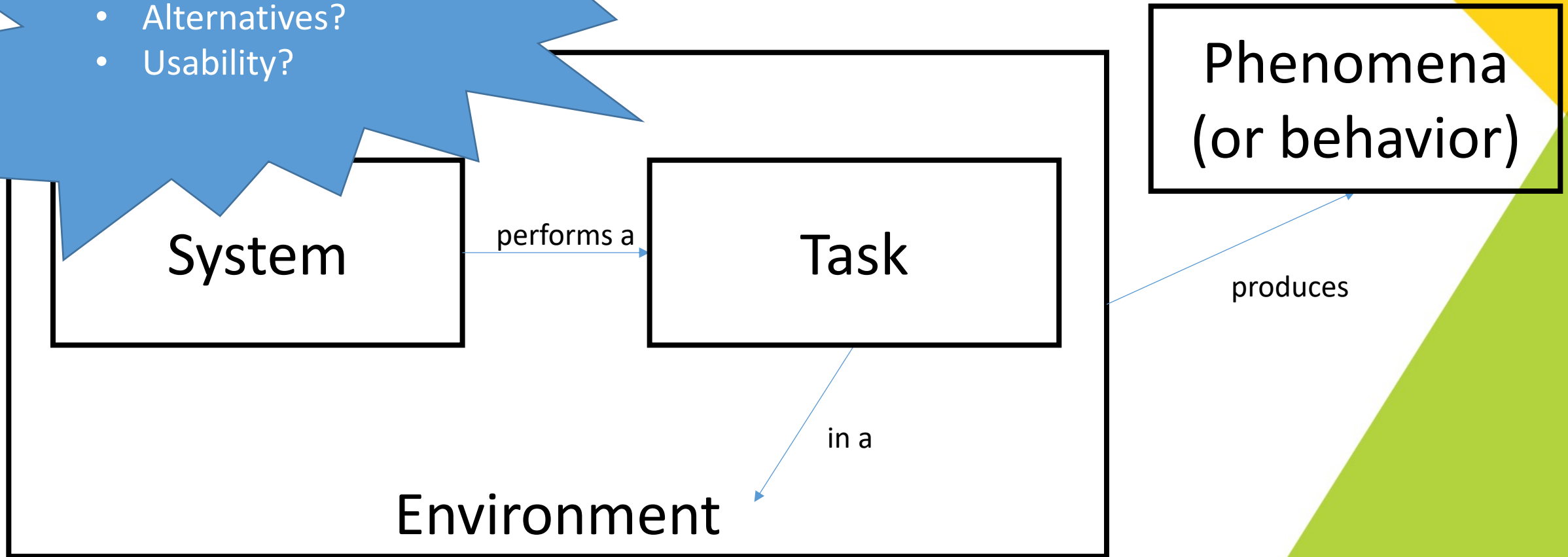
# Recall



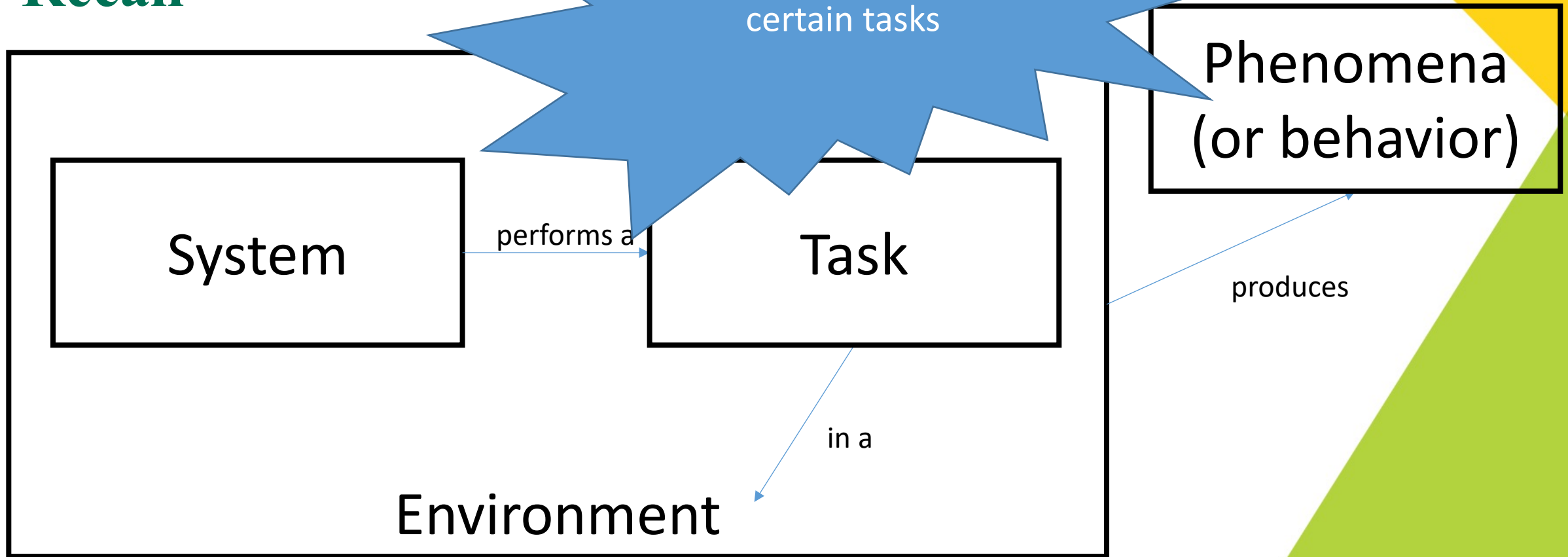
Funding to build  
systems: engineering  
vs. science



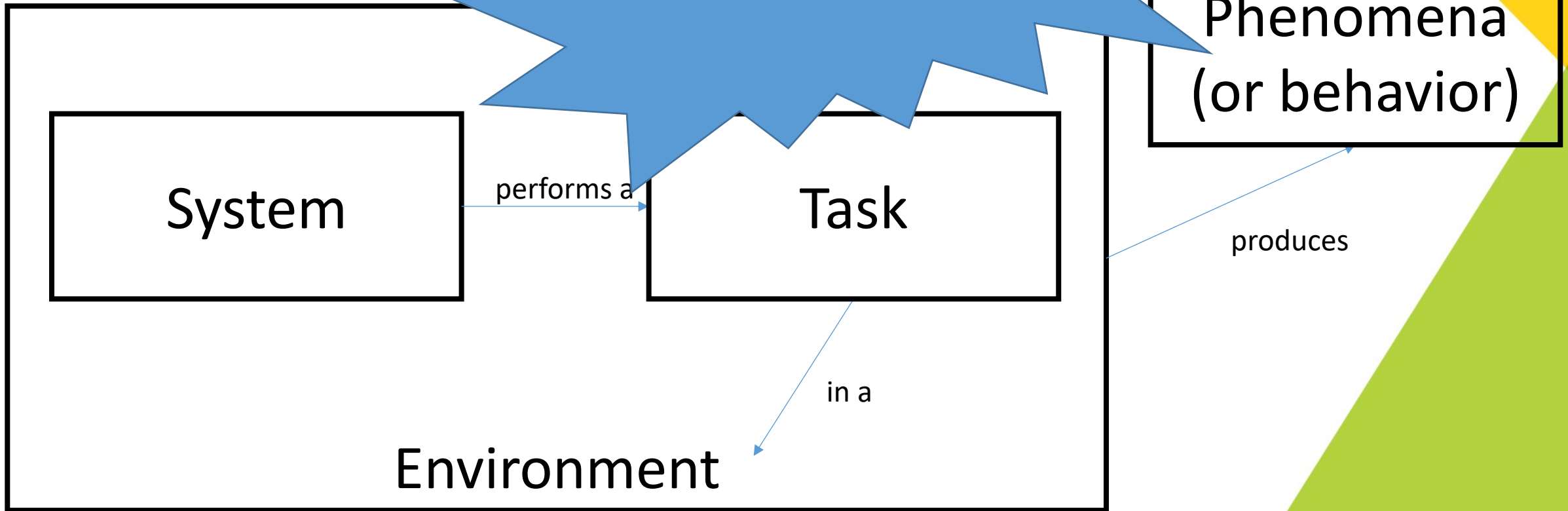
- Open source?
- Alternatives?
- Usability?



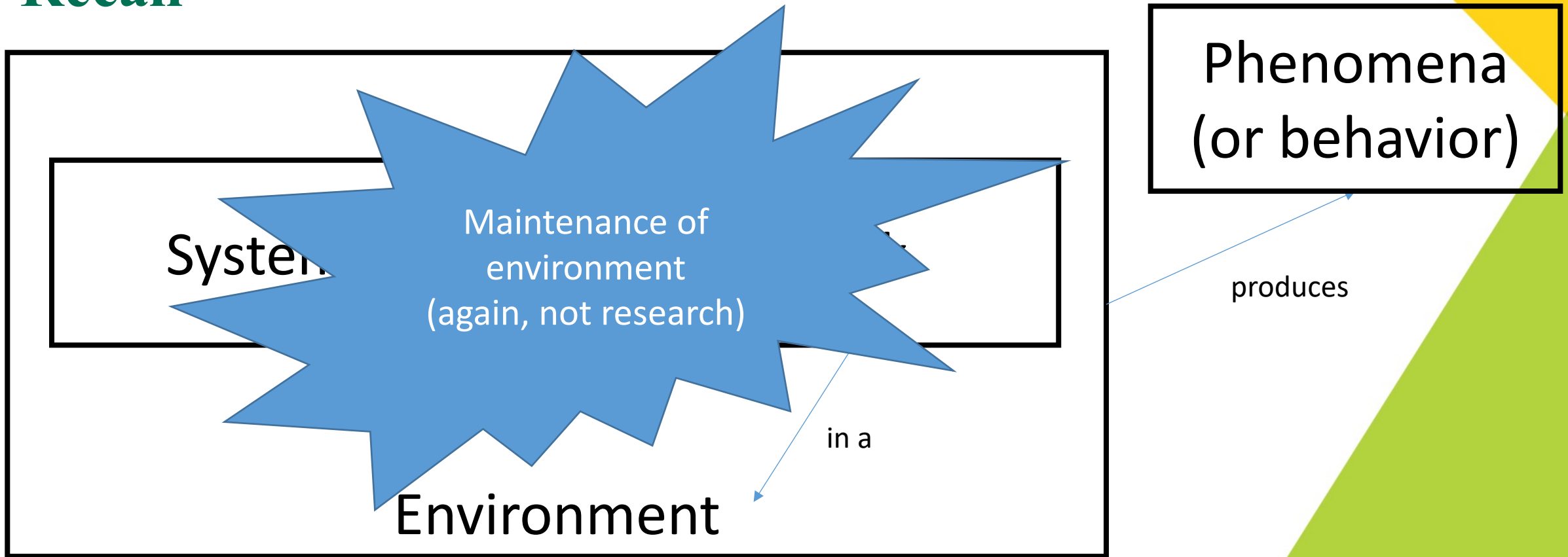
# Recall



## Recall

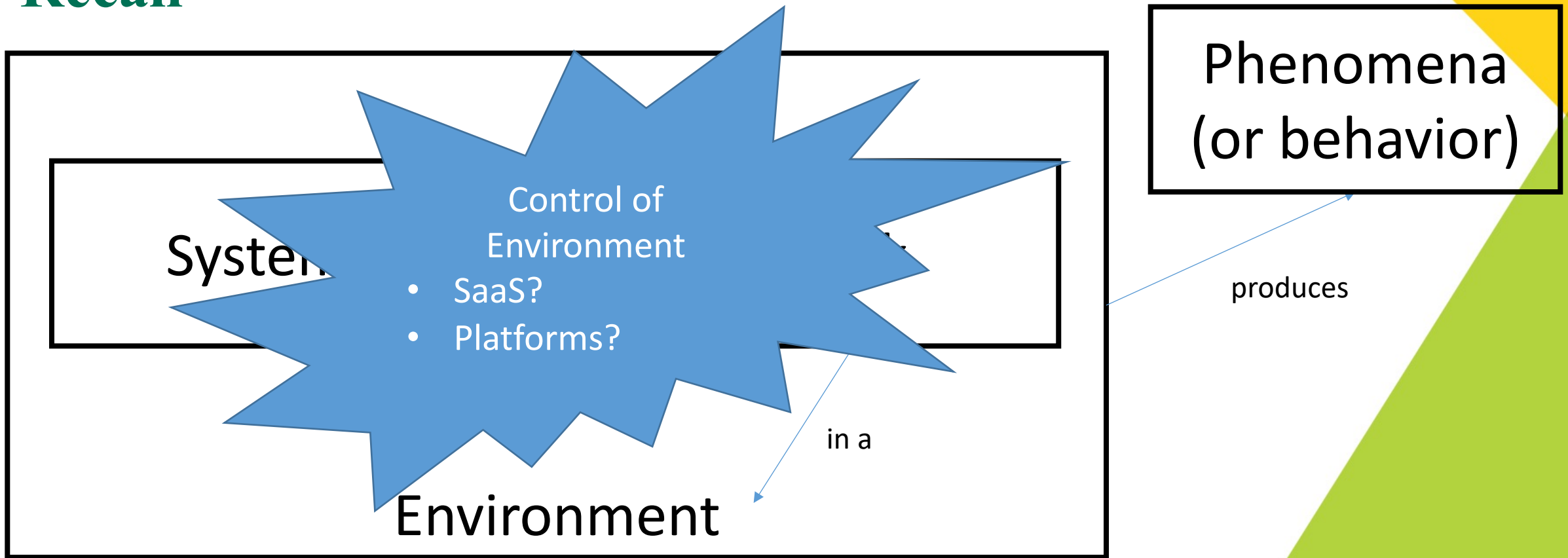


# Recall

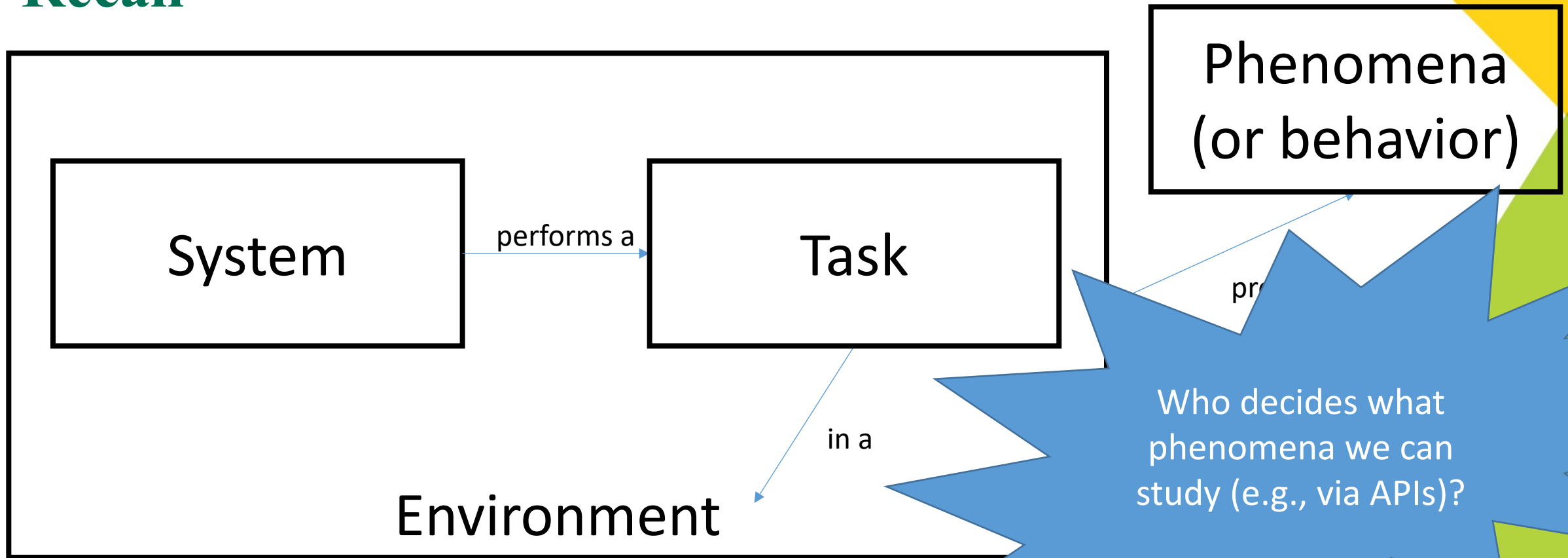




# Recall



# Recall



---

---

# Threats

- Limits on questions (e.g., phenomena)
- Non-reproducible methods (who controls the software/platforms?)
  - Is this actually bad? (see: CERN)
- No reason to register hypotheses
- Publicity machines
  - Undermines double-blind
  - Early press releases set unrealistic expectations
  - Problems with communicating with the public

---

---

# Models of Science in Industry: How much of a wall?

- **Bell Labs** – fairly strong division between basic and applied research
  - Today: similar to government contractors (MITRE, Leidos, BBN, SRI, etc.)
- **MSR** – pre-2014ish: strong division (esp. for RiSE)
  - Today: MSRNext more integrated
- **Google** – pre-2014ish: strongly integrated with applications
  - Today: Google Brain, DeepMind less integrated
- **Facebook** – tends to mimic Google

---

---

## Reflect

**Can you do good science in an industrial setting?**

- What safeguards exist?

**Can you do good science in a for-profit context?**

- What are the incentives?

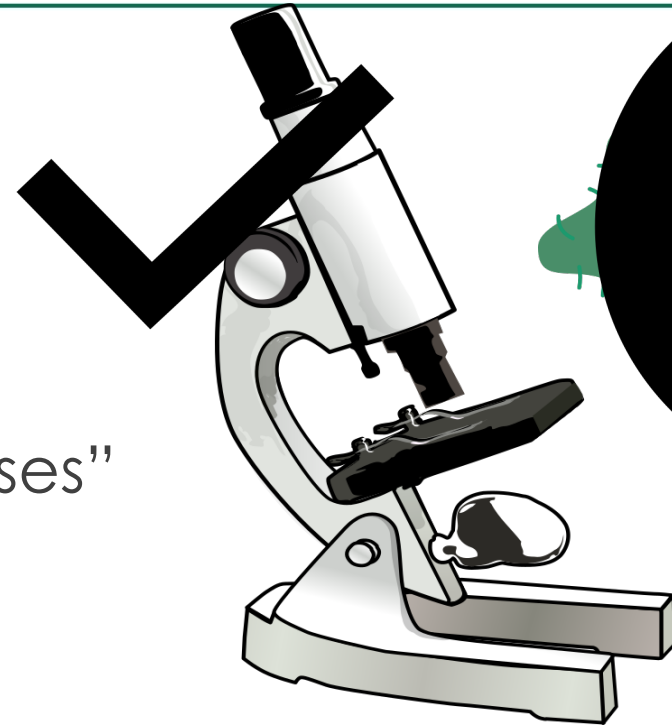
# Building Tools vs. Governing Usage

---

---

## Reminder: about studying tools

- KDD = “Knowledge Discovery in Databases”
- KDD subsumed by “data science”
- *I will **not** be teaching data analytics, **data science**, nor data mining in this course. **Instead** we will focus on **tool support** for these tasks and discuss how to design and augment existing systems, specifically for **data collection** tasks.*



---

---

Reminder: about studying tools

**Current discourse around industrial research focuses on phenomena.**

- KDD = “Knowledge Discovery in Databases”
- KDD subsumed by “data science”

• I will **not** be teaching data analytics, **data science**, nor data mining in

**Less discourse around tools.**

this course. I will be discussing how to design and augment existing systems, specifically for

**data collection tasks.**





---

---

# Tension: Tools vs. Frameworks vs. Platforms

## What is a tool?

Software designed and used for a specific purpose.

## What is a framework?

Collection of tools designed and used in concert for a specific purpose.

## What is a platform?

Complete SW/HW foundation designed and used for enabling behavior.



Scale

Complexity?

Uncertainty?

# Theme: “correctness”

## Tension: Tools vs. Frameworks vs. Platforms

- Verification and Validation
- Methods:
  - Formalisms
  - Empirical analyses
  - Case studies
- Recall: *correctness is something author/developer decides*

Scale

Complexity?

Uncertainty?



---

---

# Tension: Tools vs. Frameworks vs. Platforms

What is a tool

So

What is

Conc

W

Complete SW foundation *designed and used*

Are developers responsible for how their tools are used?

Should we update notions of “correctness” (at least, validation) in light of use?

Scale

Complexity?

ity?

---

---

## Where to draw the line?

**NRA model:** “Guns don’t kill people, people kill people.”

*Software tools: facial recognition, deep learning, formal methods?*

**Civil engineering model:** Roads can be used for many purposes.

*Software platforms: .NET platform, AWS, AMT and Facebook*

*Does software maintenance and evolution change the calculus?*

**Disclaimer: I am not a philosopher, lawyer, or ethicist**

---

---

# Governance: how control platforms-as-tools

CivilServant thesis: software platforms + users function as “digital institutions”

- Need: policy evaluation/program analysis
  - Terms from social science
  - Not to be confused with RL (policy evaluation) or PL (program analysis)

Can also think about governance as an evolving rules-based system to draw the  
line between platform and tool