Selecting Projects

CSCI 8901:
Research & Evaluation Methods

Prof. Tim Wood
GWU

These slides are based on the Research Methods course by David Jensen, UMass

Selecting a Problem

"A successful person isn't necessarily better than her less successful peers at solving problems; her pattern-recognition facilities have just learned what problems are worth solving."

– Ray Kurzweil

How to Recognize Patterns?

With training!

You need practice seeing what others currently view as important problems…

- Requires reading lots of new papers

…and you need to see which papers have had long term impact

- Requires learning the history of your field and its seminal papers

The Best CS Research…

Explores problems that:

Explain the behavior of algorithms, systems, protocols, etc

Your goal should **not** be to build the fastest X

Instead your goal should be to **understand** the principles that make X faster/better/more accurate

Good Questions

How…?

Why…?

In what situations…?

Under what conditions…?

The answers won't be yes/no!

Good Problems

“I skate to where the puck is going to be, not where it has been. ”

– Wayne Gretzky

Good Problems

"I skate to where the puck is going to be, not where it has been. "

– Wayne Gretzky

Not easy to do… but might as well try!

My NSDI 2007 paper…

Black-box and Gray-box Strategies for Virtual Machine Migration

Timothy Wood, Prashant Shenoy, Arun Venkataramani, and Mazin Yousif[†]
Univ. of Massachusetts Amherst [†]*Intel, Portland*

Abstract

Virtualization can provide significant benefits in data centers by enabling virtual machine migration to eliminate hotspots. We present Sandpiper, a system that automates the task of monitoring and detecting hotspots, determining a new mapping of physical to virtual resources and initiating the necessary migrations. Sandpiper implements a black-box approach that is

tualization of data center resources provides numerous benefits. It enables application isolation since malicious or greedy applications can not impact other applications co-located on the same physical server. It enables server consolidation and provides better multiplexing of data center resources across applications. Perhaps the biggest advantage of employing virtualization is the ability to flexibly remap physical resources to virtual servers in

Proposed techniques for more efficiently and dynamically managing servers in a data center

- Some good ideas
- Some pretty bad ideas (clear issues for how it modeled the combined load of different resource types

My NSDI 2007 paper…

Black-box and Gray-box Strategies for Virtual Machine Migration

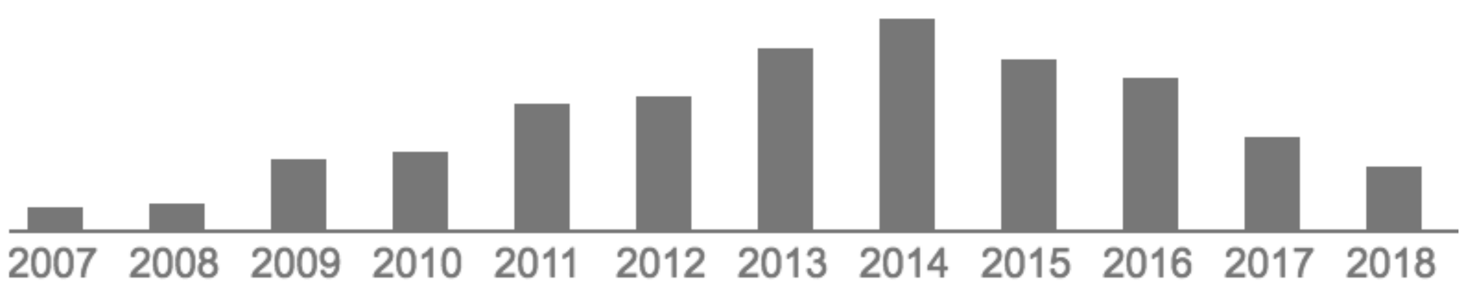
Timothy Wood, Prashant Shenoy, Arun Venkataramani, and Mazin Yousif[†]
Univ. of Massachusetts Amherst [†]*Intel, Portland*

Abstract

Virtualization can provide significant benefits in data centers by enabling virtual machine migration to eliminate hotspots. We present Sandpiper, a system that automates the task of monitoring and detecting hotspots, determining a new mapping of physical to virtual resources and initiating the necessary migrations. Sandpiper implements a black-box approach that is

tualization of data center resources provides numerous benefits. It enables application isolation since malicious or greedy applications can not impact other applications co-located on the same physical server. It enables server consolidation and provides better multiplexing of data center resources across applications. Perhaps the biggest advantage of employing virtualization is the ability to flexibly remap physical resources to virtual servers in

Cited by 1048



Nice to get lots of citations…

My NSDI 2007 paper…

Black-box and Gray-box Strategies for Virtual Machine Migration

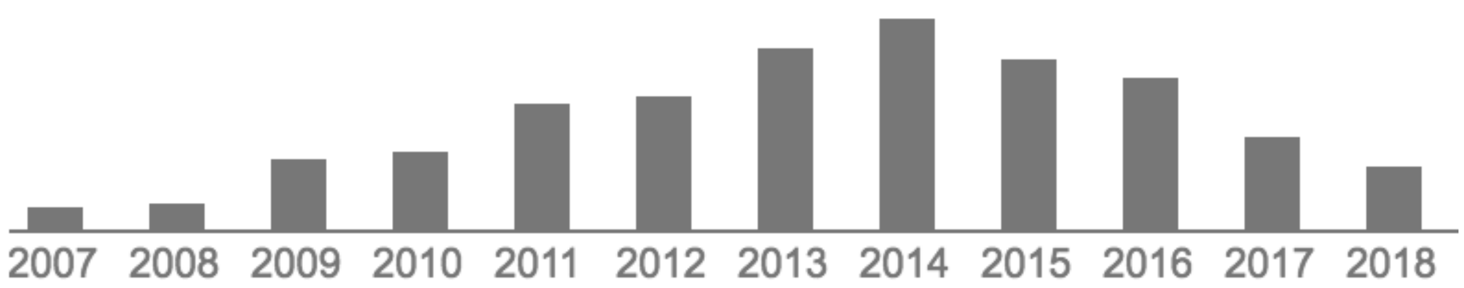
Timothy Wood, Prashant Shenoy, Arun Venkataramani, and Mazin Yousif[†]
Univ. of Massachusetts Amherst [†]*Intel, Portland*

Abstract

Virtualization can provide significant benefits in data centers by enabling virtual machine migration to eliminate hotspots. We present Sandpiper, a system that automates the task of monitoring and detecting hotspots, determining a new mapping of physical to virtual resources and initiating the necessary migrations. Sandpiper implements a black-box approach that is

tualization of data center resources provides numerous benefits. It enables application isolation since malicious or greedy applications can not impact other applications co-located on the same physical server. It enables server consolidation and provides better multiplexing of data center resources across applications. Perhaps the biggest advantage of employing virtualization is the ability to flexibly remap physical resources to virtual servers in

Cited by 1048



But why are some people still doing the same thing 10+ years later?

New and Old

Of course you should read and cite “old” work!

- In CS, we often forget about things older than ~7 years

But you should do more than just iterate on it!

Think about what has changed…

- Underlying technologies
- Workloads of users / data
- Advances in algorithms or modeling techniques

Use these advances to make a bigger change

Reading 1

A Guide to Increased Creativity in Research — Inspiration or Perspiration?

by C. Loehle, in Bioscience February 1990

A guide to increased creativity in research—inspiration or perspiration?

There are four requirements for a successful career in science: knowledge, technical skill, communication, and originality or creativity. Many succeed with largely the first three. Those who are meticulous and skilled can make a considerable name by doing the critical experiments that test someone else's ideas or by measuring something more accurately than anyone else. But in such areas of science as biology, anthropology, medicine, and theoretical physics, more creativity is needed because phenomena are complex and multivariate.

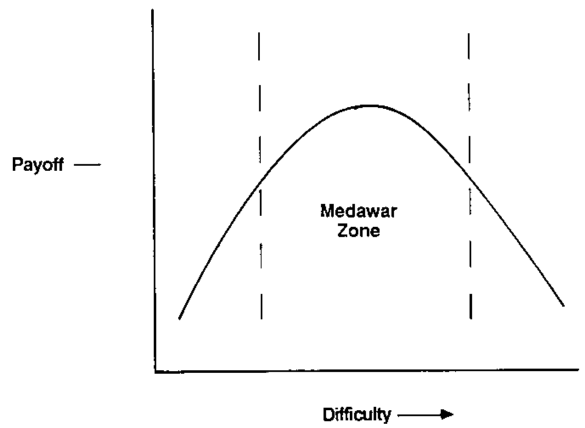
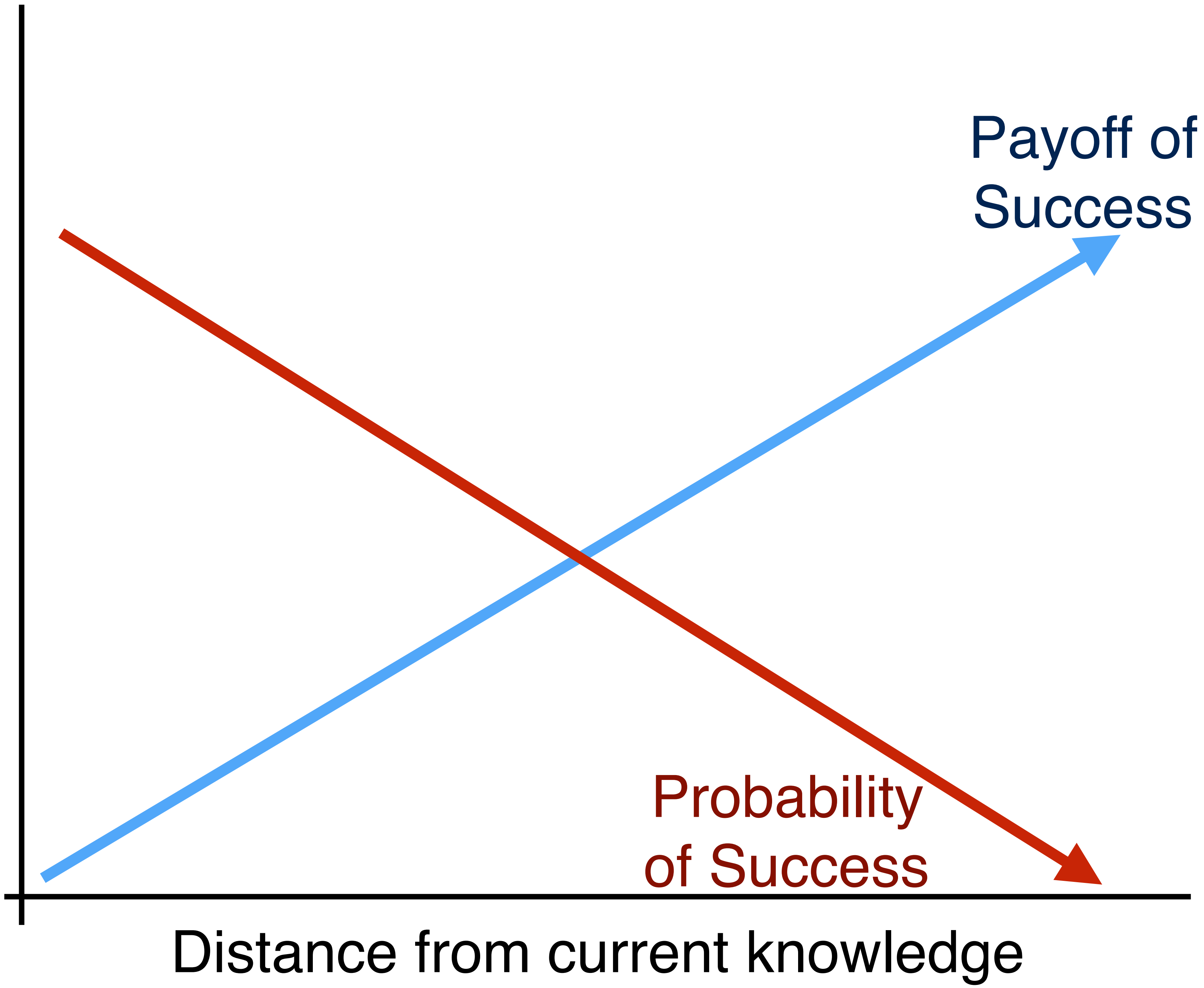


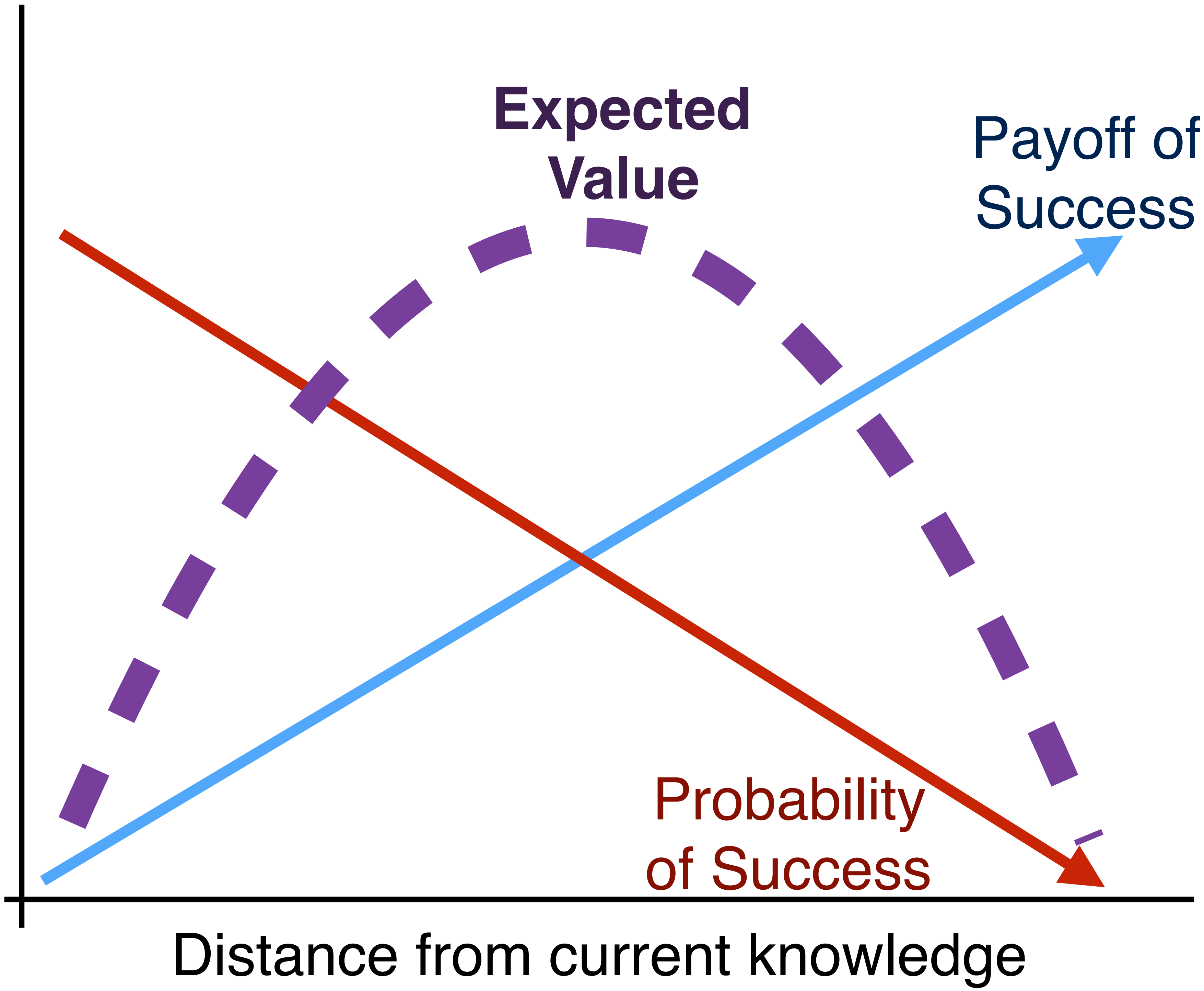
Figure 1. Relationship between degree of difficulty and payoff from solving a problem. Solving problems that are too easy

and work intensely. After you have finished writing your paper, you can go back and remove the comments about what an imbecile the other person is. The effort to refute someone can even lead to evidence supporting them or to a different topic altogether. Intensive rivalries, as in the race to discover DNA (Watson 1968), can also provide this essential intensity. Thus whereas the finished product may appear dispassionate, truly creative work is often driven by strong passions.

Medawar Zone



Medawar Zone



Other Useful Tips?

Don't be an expert

Develop courage

Don't read the literature?

Pick a fight / where is the smoke?

Writing as brain dumps

Finding the Frontier

New results that don't have a clear explanation yet

Opportunities provided by…

- Changing hardware
- New algorithms
- New modeling frameworks

The convergence of two fields

New applications of your field to another

"Good research is done with a shovel, not with tweezers... You should find an area where you can get a lot out of it fast."

– Roger Needham

Pick the right problem… for you!

What are your strengths?

- Algorithmic, implementation, experimental, analytic, etc

Why will you be able to solve this problem more effectively than most other people?

Pick the right problem… for you!

What are your strengths?

- Algorithmic, implementation, experimental, analytic, etc

Why will you be able to solve this problem more effectively than most other people?

Bad answers:

- “I’m smarter than most people!” (Sorry, but this is unlikely)
- “I will work harder and longer!” (Unlikely and can hurt the other parts of your life)

“Always have a secret weapon — the biggest computer, a problem imported from another field that others haven't heard of yet, a fact you stumbled on by being curious about everything, a friend who is smarter than you are. Anything, in short, that will give you an unfair advantage in getting there first.

– Herbert Simon

A necessary ingredient…

Pick a project that matches your **personal interests**

Research takes a long time

- You need to stay motivated and interested in the topic!

Maintaining your enthusiasm is important

- For yourself
- For convincing others to join / help you

Typical Research Timeline

Where does idea come from?

What do you do first?

How long to build and test your solution?

How long to write a paper?

How long until paper is reviewed?

Typical Research Timeline

Where does idea come from?

- From advisor, from literature

What do you do first?

- read a lot about the problem, learn about tools/implementation aspects, write list of contributions, talk to your advisor, do motivational experiments

How long to build and test your solution?

- 1-2 years

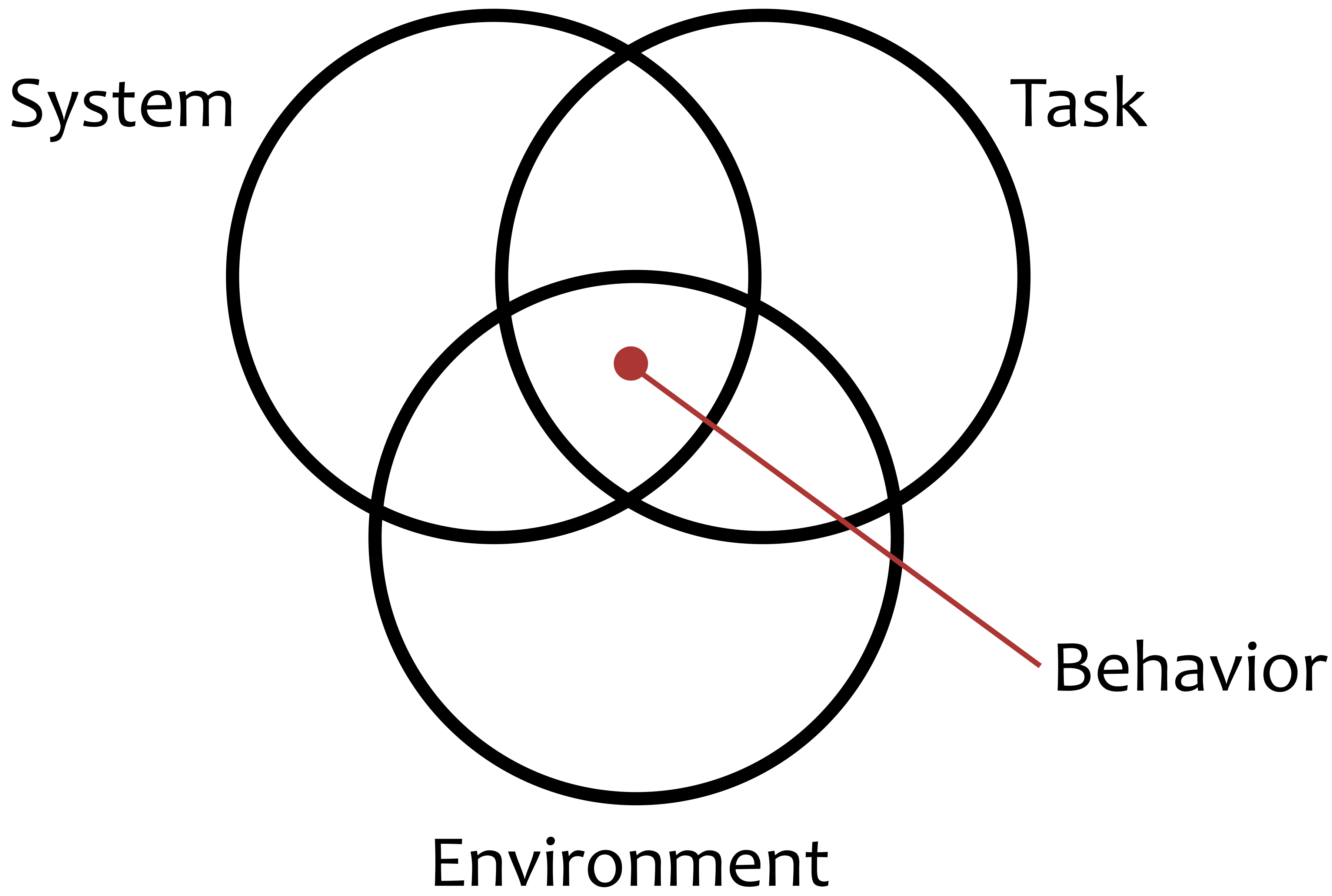
How long to write a paper?

- 2-3 weeks often last minute. Better to start some pieces earlier

How long until paper is reviewed?

- 4 months, might take multiple submissions to be accepted

Research Problem Framework



From Paul R. Cohen, Empirical Methods for Artificial Intelligence. MIT Press 1995

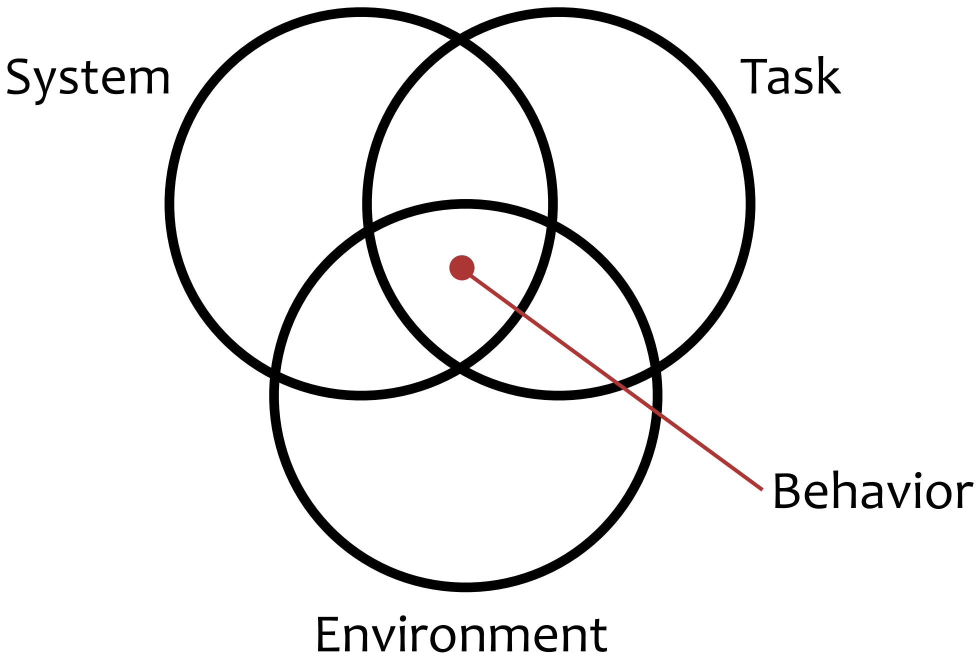
System

The thing you "built"

Could be anything:

- a new operating system
- a new ML inference algorithm
- a new NLP framework
- a wireless network management algorithm

This part of the research is fully **under your control**

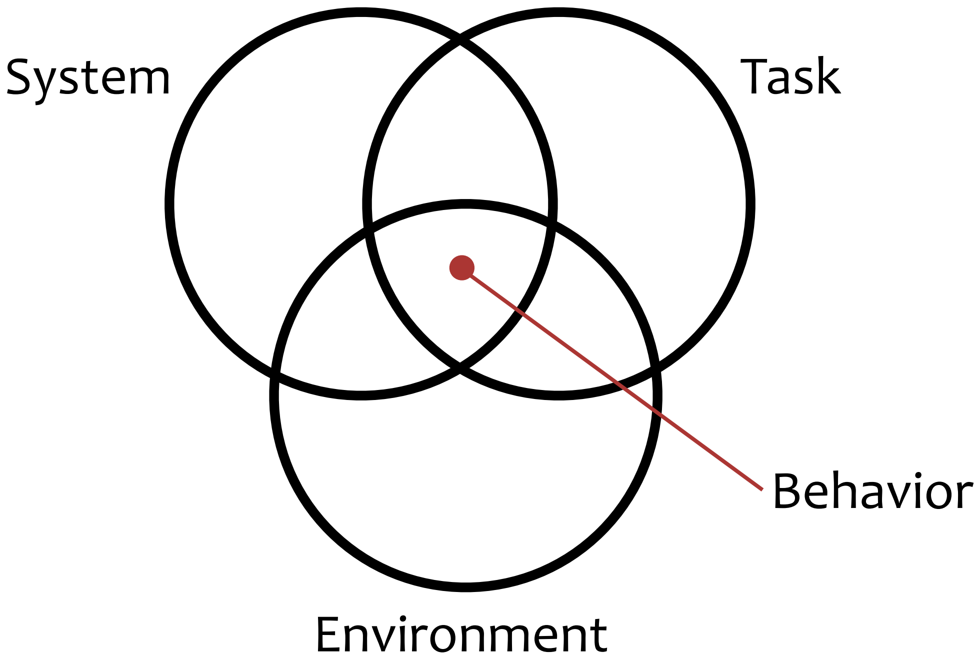


Task

The "work" to be done

This is an input to your system:

- Incoming requests to a web server
- The type of query to a DB
- The input data be analyzed by your algorithm

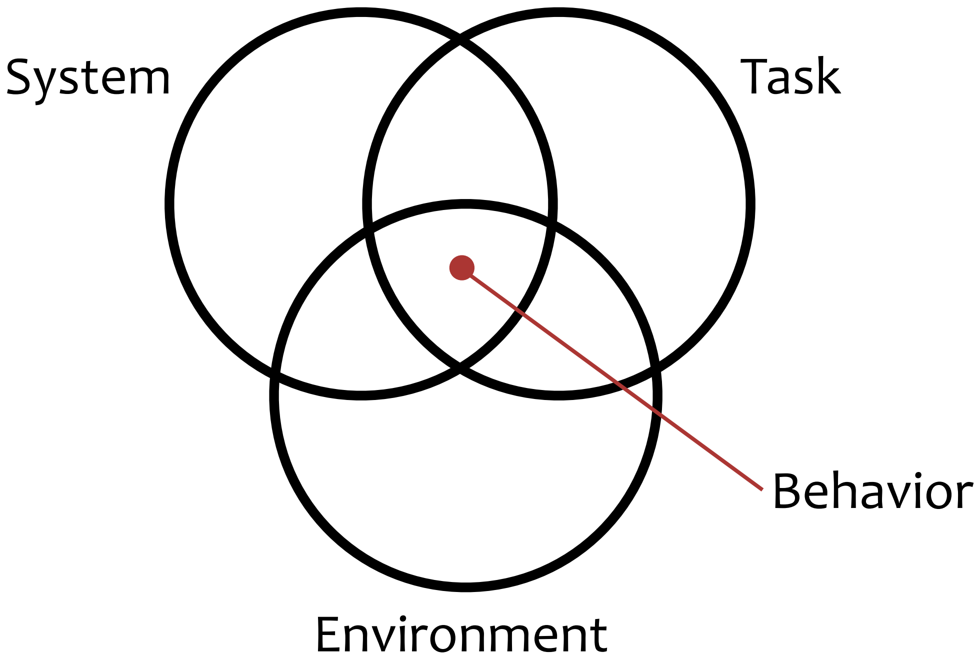


This is **under the control of the user** of your system

Environment

The surroundings that affect both the task and the system

- Available CPU and memory
- The network characteristics

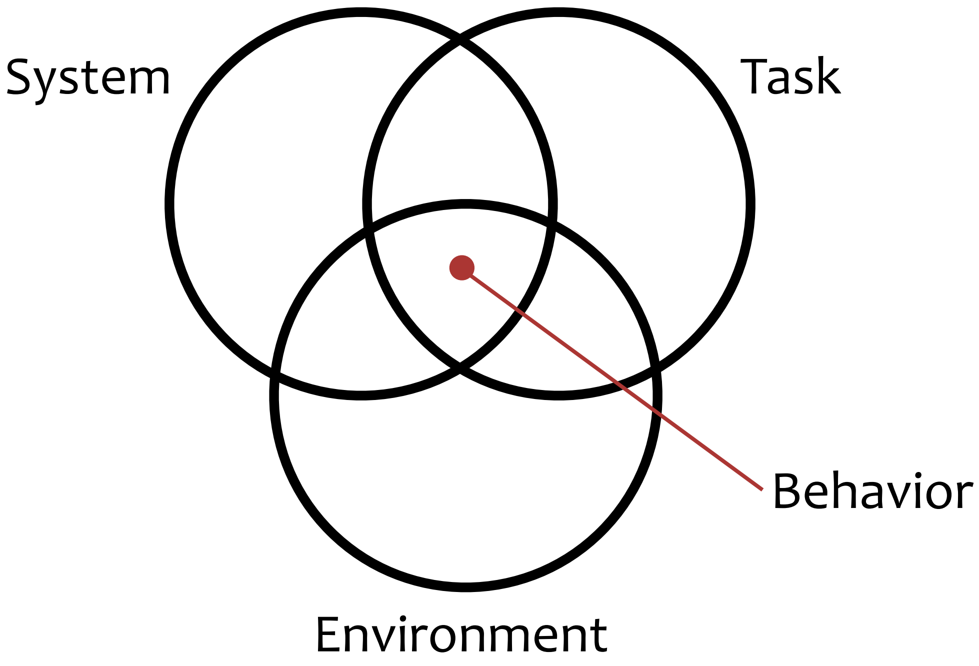


This is **can't be controlled by anyone**

Behavior

How does the system behave for a particular task and environment?

- Your class project should focus on this type of behavior!



You should evaluate how the

- system, task, and environment

affect the behavior

Acknowledgements

Much of the slide content, and almost all of the amazing quotations, are derived from the *Research Methods for Empirical Computer Science* course taught by **David Jensen**

- http://dx.doi.org/11084/10002
- https://people.cs.umass.edu/~jensen/courses/index.html
- https://people.cs.umass.edu/~jensen

- Many thanks for allowing me to make use of his materials!