# Picking Research Projects

Interesting CS research questions

- Explore or explain the behavior of algorithms, systems, protocols, and other computational artifacts.
- Typical form of research questions:
  - "Why…"
  - "How…"
  - "Under what circumstances…"
  - "What are the necessary and sufficient conditions for…"
- Answers: not yes/no, but paragraphs, mathematical or statistical models, simulations, etc.

Framework for research problems
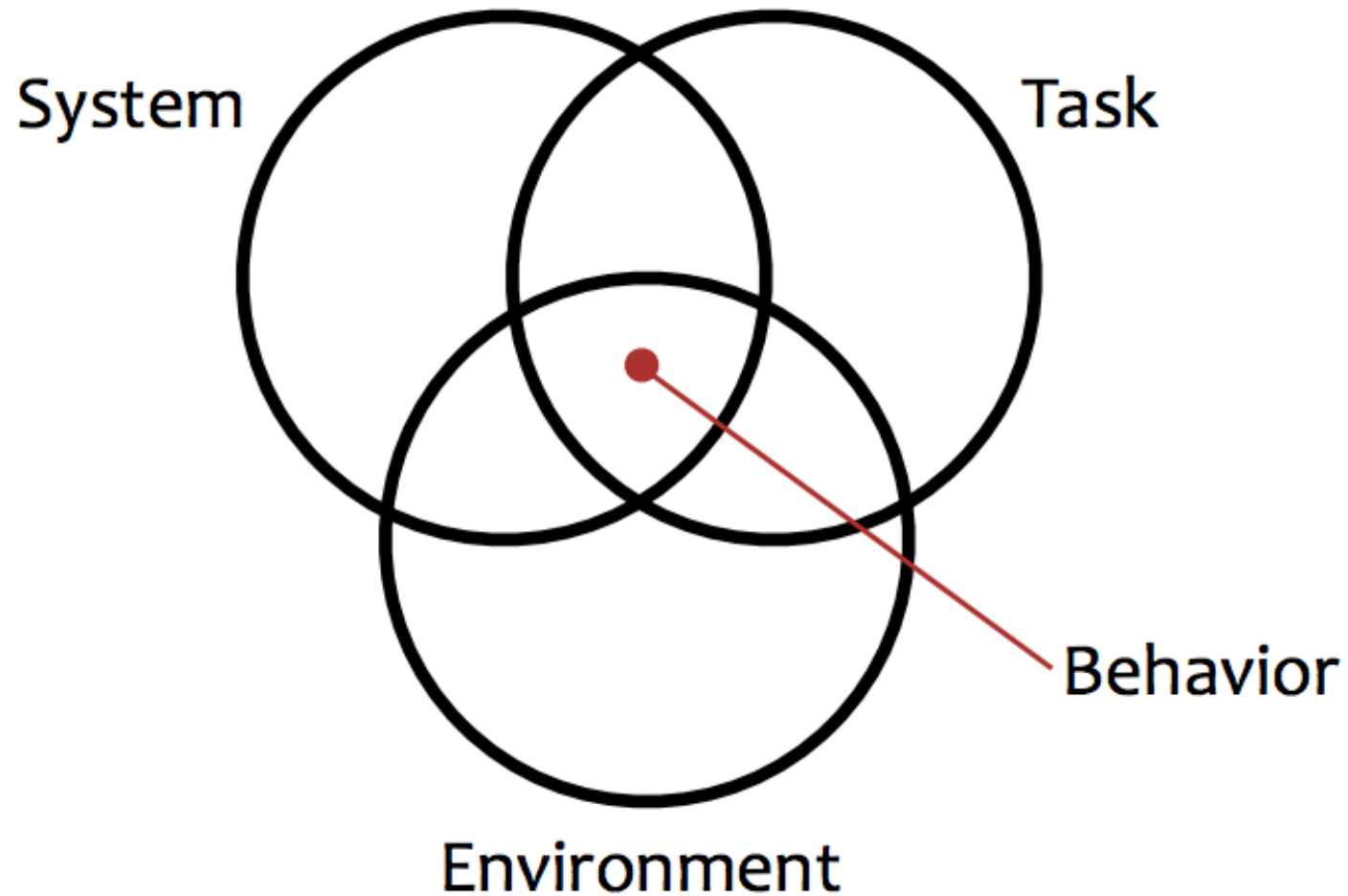
Framework for research problems

- System: aspects influenced by a system designer.
  - Specific algorithm used, system architecture, data structures, parameter settings, etc.
- Task: aspects influenced by a prospective user.
  - Specific queries, requests, input data, etc.
- Environment: aspects influenced by neither a designer or a user.
  - Network environment, availability cycles or memory, etc.
- Behavior: performance of the task by the system within the environment.

- What are aspects of the system?
  - PageRank, server farms, indices
- Task?
  - queries, preferences, language
- Environment?
  - document distribution, network load
- Behavior?
  - Retrieval performance

Factors to consider

- Importance
  - How important is the research topic within the larger research and application community?

- State of knowledge
  - What do we know already? What is the position of the research with respect to "the frontier?"

- Unique competence
  - Are you uniquely qualified to address this research? What is your "secret weapon?"

- Interest
  - How much does this research problem interest you personally? Do you have a passion for this problem?

Assessing importance

- Audience
  - Who will care about the answer?

- Impact
  - Will different answers change what research is done next? ...what is done by practitioners?

- Longevity
  - How long will the answer be relevant and important?

How many CS talks begin

- Graph of rapid growth in…
  - Processing power (most common)
  - Heat generation
  - Relative size or cost of caches
  - Size of the web
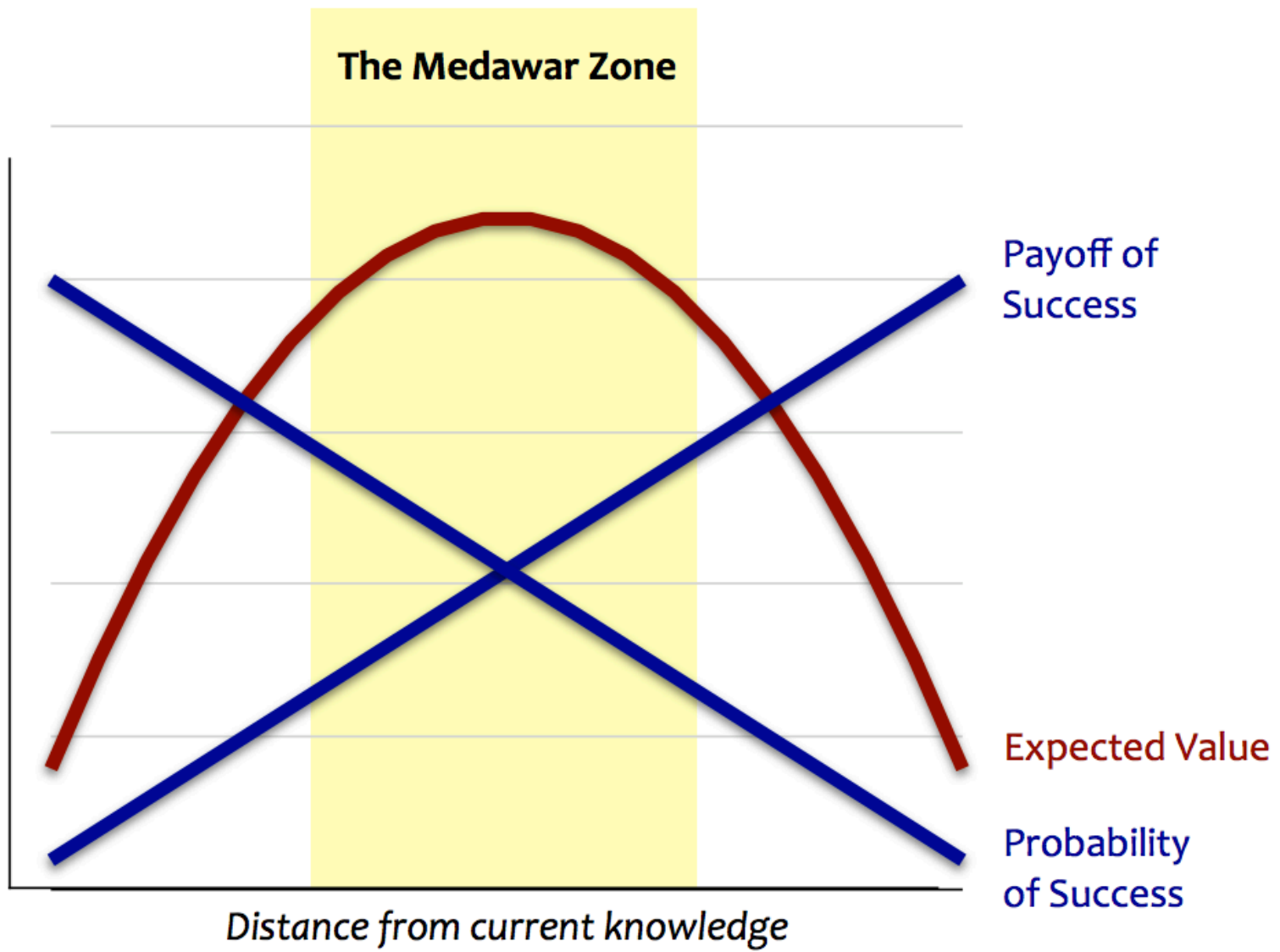 - Installed base of specific devices
  - …
- Why?

"The Frontier"

- Important work typically takes place at the *frontier* of a field.

- You want your research findings to be relevant when they are published, not just now.

- The frontier in CS has two components:
  - Changing face of CS knowledge (a problem common to all fields of science)
  - Changing environment of computer technology and applications (a particularly challenging element of computer science)

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

– Wayne Gretzky

The Medawar Zone

Payoff of Success

Expected Value

Probability of Success

Distance from current knowledge

Methods for identifying frontiers

- New observations of findings that need explanation.
- New opportunities provided by instruments, methods of study, or theoretical frameworks.
- Converging lines of research that combine to provide new opportunities.
- Emerging needs from outside the field (e.g., web search, bioinformatics, ecological monitoring, autonomous vehicles).
- New questions or conjectures by researchers with good track records.

"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

Unique competence

- What is it that makes you better able to address this research question than others?
- What is your "secret weapon?"
- Poor answers
  - *"I am smarter than other people."*
    (rarely true)
  - *"I will work harder, longer, or faster."*
    (There is always someone who can to those things better than you, and that weapon can cause unacceptable losses: harm to family, friends, and personal happiness.)

"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

Personal interest

- Focus your work on areas of personal interest.
- You will work harder, smarter, and more creatively.
- Others will sense your strong interest and want to work with you, support your work, and credit you with innovations.
- Personal interest can be fostered and destroyed. Track what affects your interest and use that knowledge.
  - One widely shared factor: *speed of progress*.

Some properties of a good senior seminar project

- Located at the frontier
  - Identifies an unexplored (or underexplored) question.
  - Experts have significant questions about the outcome.
- Involves experiments.
- Identifies independent and dependent *variables*.
- Involves understanding behavior, not just design.
- Practical issues:
  - May I build something? Yes. But you must also experiment on it.
  - Better: look for something already built, or an existing algorithm, and figure out how and why it works.

Ideas

- Build a system (or systems) and study it/them.
  - Music recommendation algorithm, new AI for intelligent NPCs in games, web-based system for housing selection, virtual environments (Dr. Sanders)
- Or use an existing system (code is easy to find or implement)
- What happens when you vary the system, task, and/or environment?
  - What behavior emerges? What can you learn? Can you make improvements?

How to succeed in research

- Be careful of framing your work in terms of "building something to do X."
  - What if X is really hard or impossible?

- What lots of people think of as "doing CS" is really *building the infrastructure* for doing CS.
  - Examples: building new compilers, garbage collectors, networking protocols, machine learning algorithms...

Research questions

- Identify a research question about which a hypothesis can be formulated.

- Typically about algorithms (systems), task, or environments.

- Questions about
  - individual elements (existence proofs)
  - how changes in one element affect another
  - comparisons of two or more elements holding others constant

Heilmeier Questions

- What are you trying to do? Articulate your objectives using absolutely no jargon.
- How is it done today, and what are the limits of current practice?
- What's new in your approach and why do you think it will be successful?
- Who cares?
- If you're successful, what difference will it make?
- What are the risks and the payoffs?
- How much will it cost?
- How long will it take?
- What are the midterm and final "exams" to check for success?