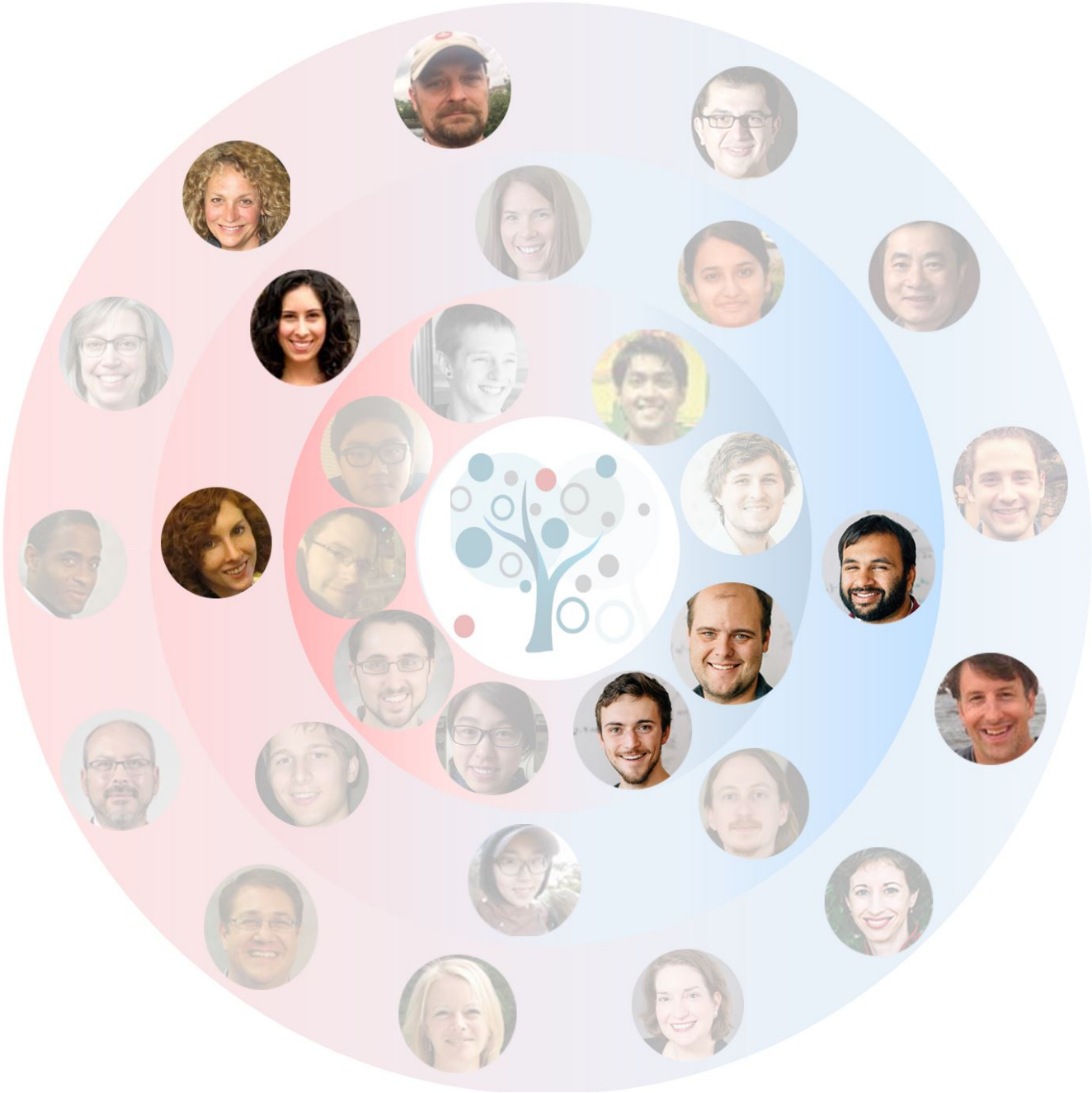
NEXT: Crowdsourcing, machine learning and cartoons

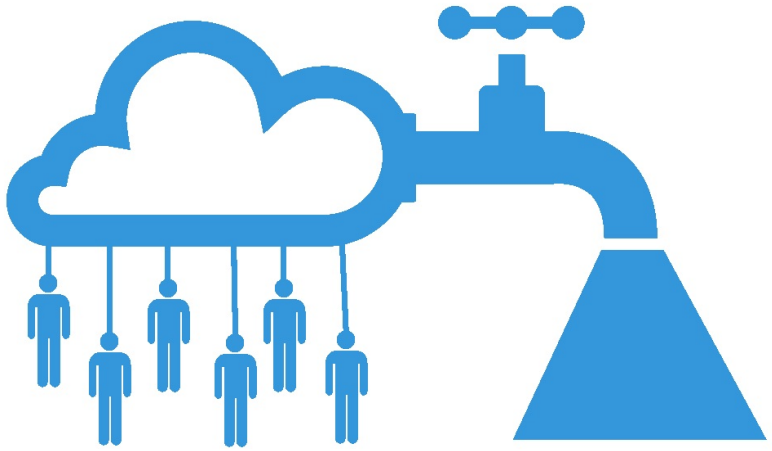
Scott Sievert
UW–Madison ECE



Joint work with Lalit Jain, Daniel Ross,
April Murphy and Rob Nowak

Problem

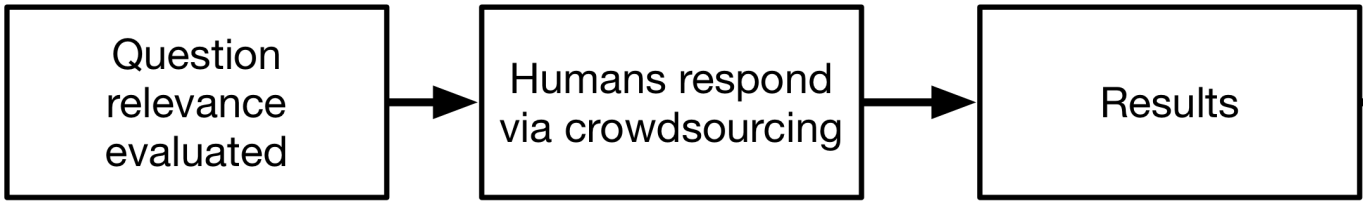
Data collection with
crowdsourcing can be expensive



Goal: achieve goal with minimal responses

One solution

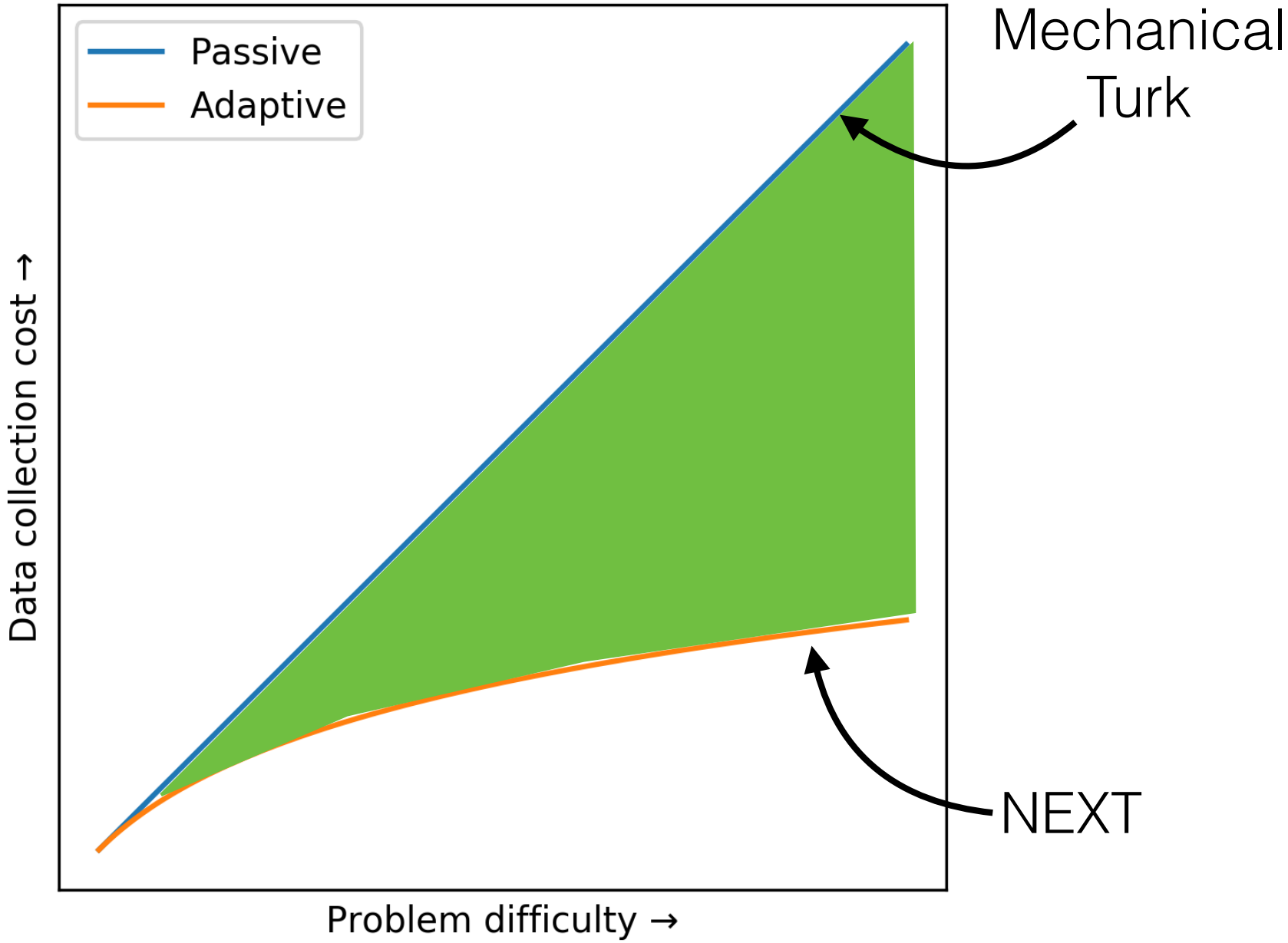
Existing crowdsourcing systems are *passive*



Adapting to previous responses requires fewer data

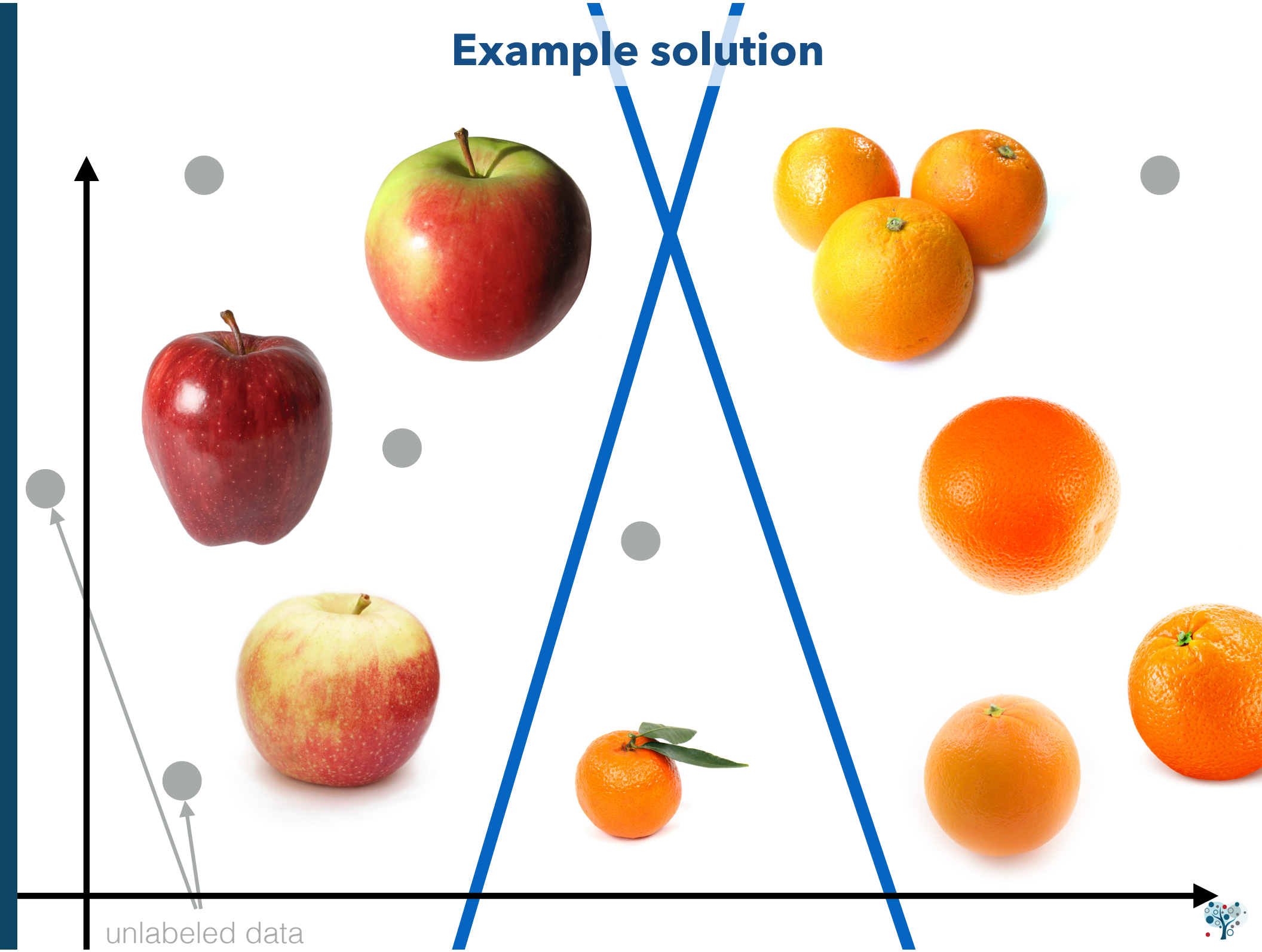
Goal: adapt to previously collected responses

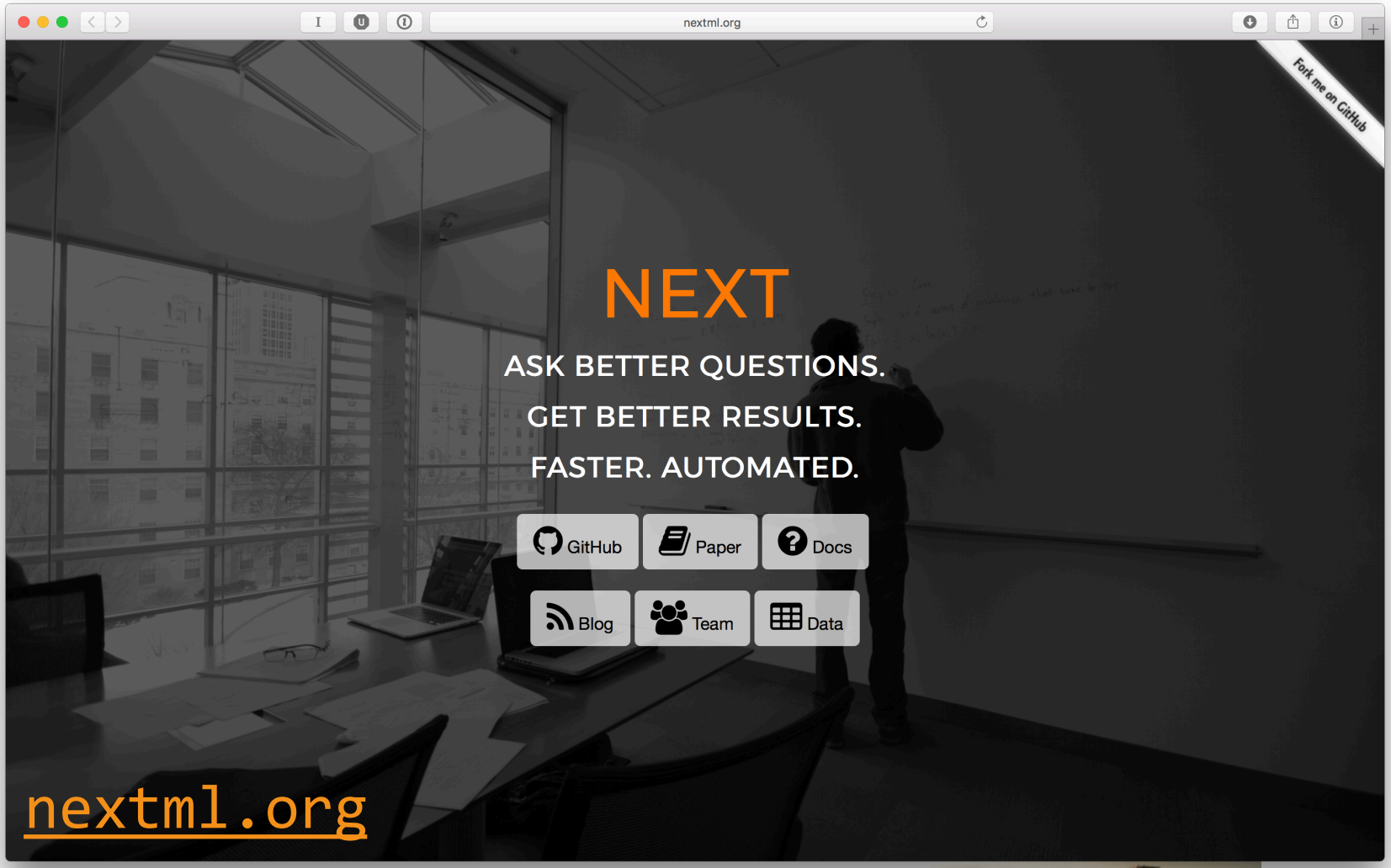
Benefits



Adaptive sampling can have large benefits

Example solution





Lalit Jain

Daniel Ross

Rob Nowak

Kevin Jamieson

Homepage: http://nextml.org
Source: https://github.com/nextml/NEXT
Documentation: https://github.com/nextml/NEXT/wiki

NSF

powered by amazon web services

WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

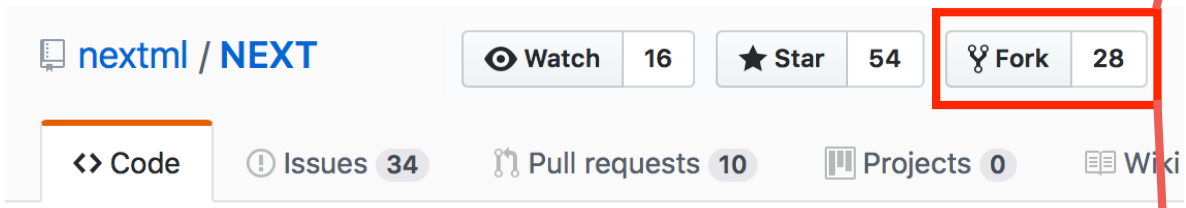
Sandia National Laboratories

amplab

AIR FORCE RESEARCH LABORATORY

LUCID
learning.understanding.cognition.intelligence.data science

NEXT users

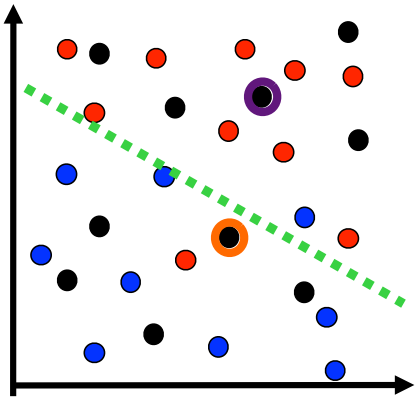


Theory

Practice

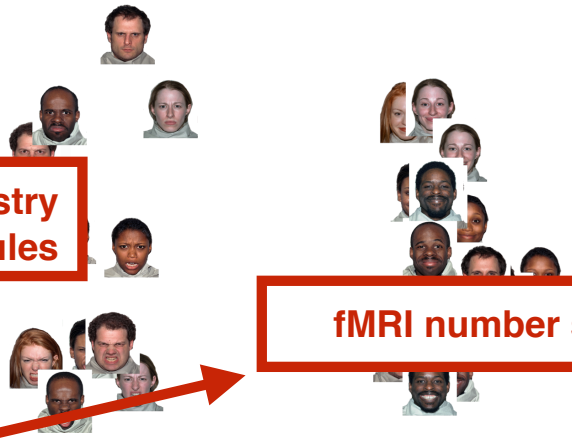
ML Researchers

Air Force Research Lab uses NEXT for active image classification.



Experimentalists

UW Psychology uses NEXT to find the best algorithms for adaptive data collection in cognitive science.



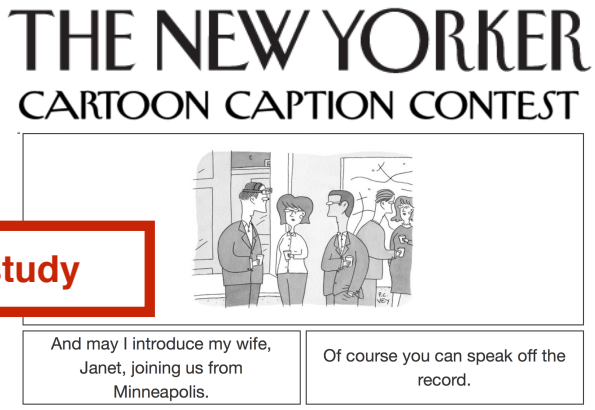
Chemistry molecules

fMRI number study

To be presented

Practitioners

The New Yorker uses NEXT to crowd-source the weekly cartoon caption contest.



- nextml / NEXT
  - aashish24 / NEXT
  - abiswas3 / NEXT
  - alphaprime / NEXT
  - aniruddhajb / NEXT
  - AvinWangZH / NEX
  - ayonsn017 / NEXT
  - caomw / NEXT
  - connectthefuture /
  - crcox / NEXT
  - dconathan / NEXT
    - robinsonkwame
  - jattenberg / NEXT
  - jimwmg / NEXT
  - justicelee / NEXT
  - juthawong / NEXT
  - liamim / NEXT
  - mllewis / NEXT
  - NandanaSengupta
  - naveendennis / NE
  - pedmiston / NEXT
  - samim23 / NEXT
  - stsievert / NEXT
    - BhargavaA / NE
  - suchow / NEXT
  - sumeetsk / NEXT-1
  - widoptimization-wi
  - worldbank / NEXT

Example problem

Bob Mankoff

THE NEW YORKER

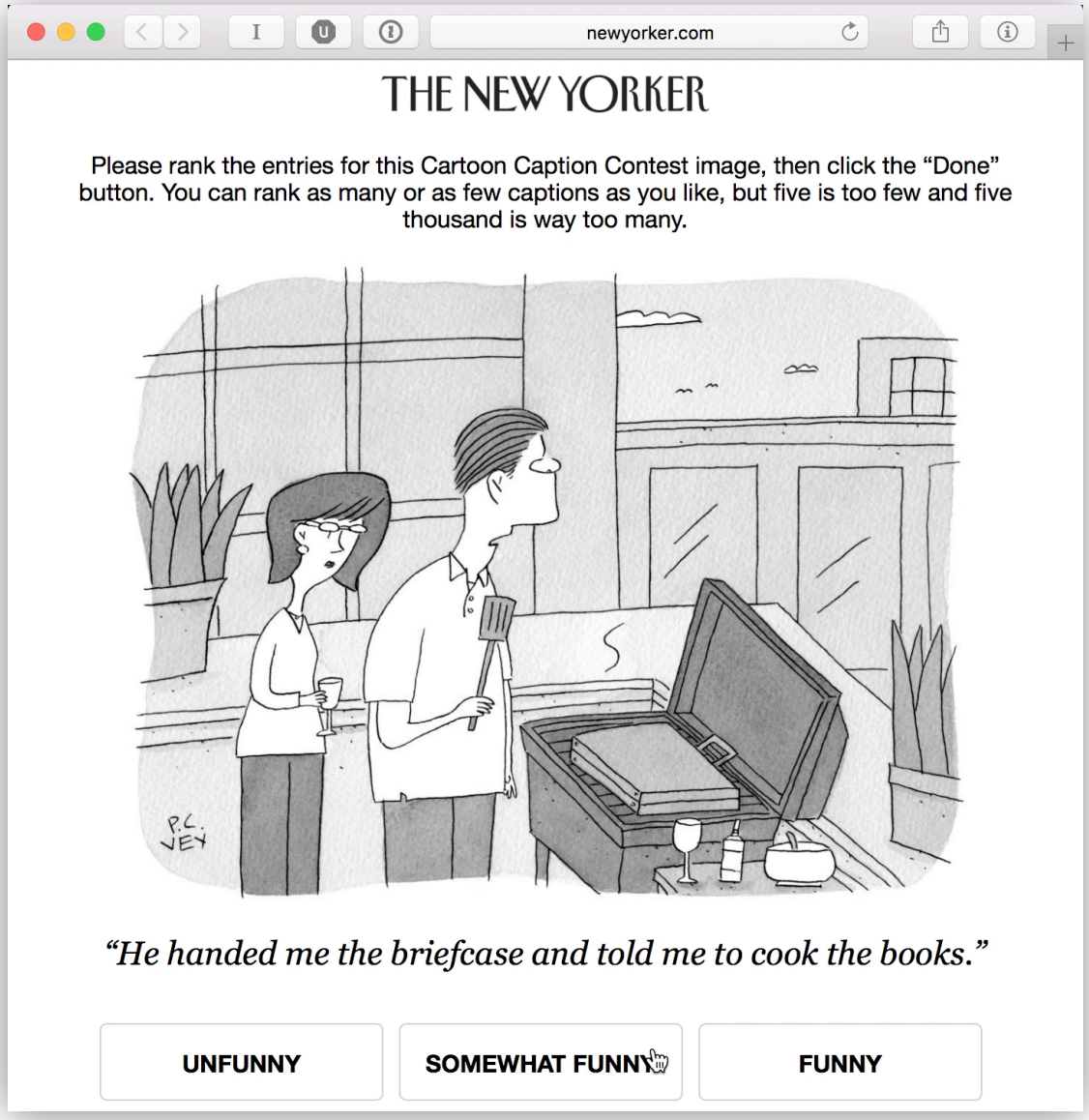
YOUR CAPTION

Enter your caption (250 characters or fewer):

The New Yorker has to find the funniest caption from ~5,000 captions

Comic by P. C. Vey

Interface



http://www.newyorker.com/cartoons/vote

http://nextml.org/captioncontest

Comic by P. C. Vey

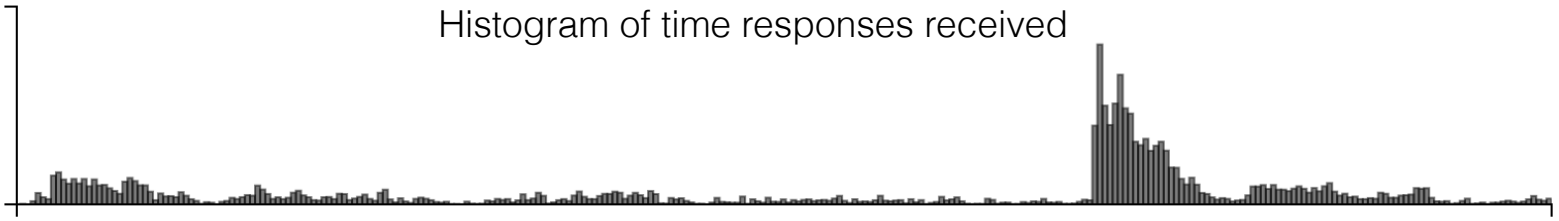
Dashboard

ec2-34-208-101-150.us-west-2.compute.amazonaws.com

NEXT Experiment Dashboard | NEXT - b85faed0ebbd9c6256bdcaa5f2951a

NEXT

EXPERIMENT SYSTEM



Rankings

KLUCB

Download CSV

Experiment Info

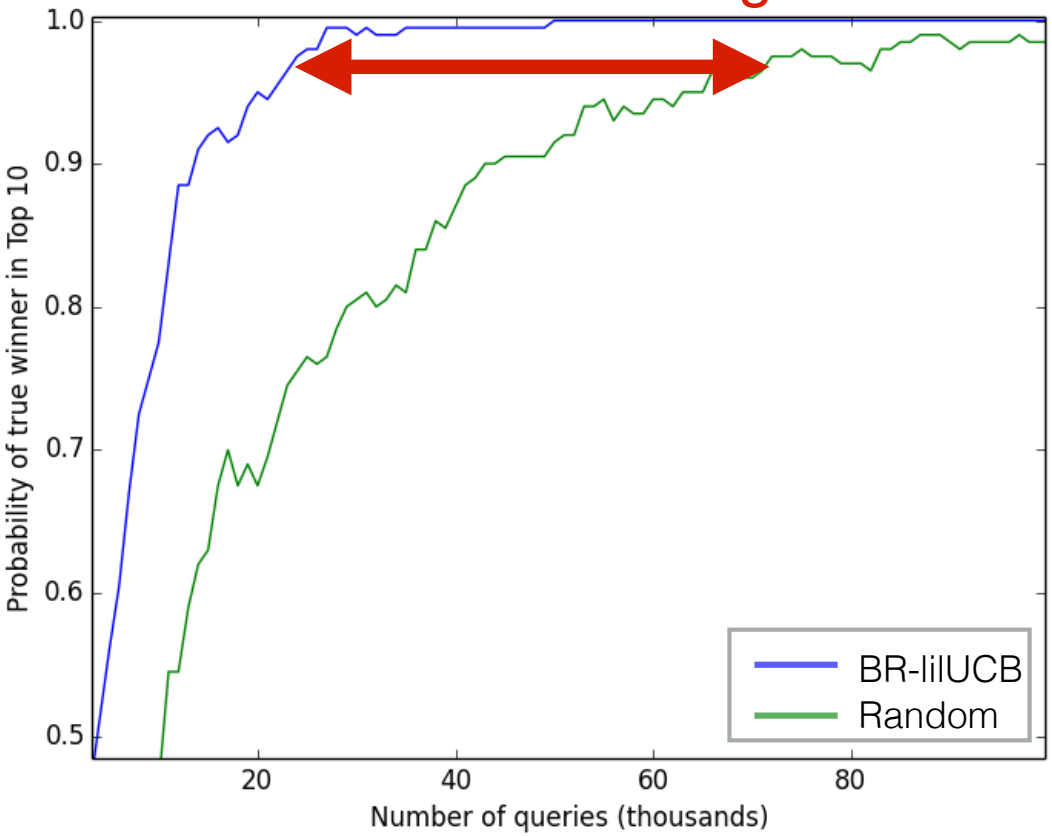
Start date: 2017-06-26 14:39:13.745959 UTC
Dashboard data generated: <1 minute ago
Number of participants: 14111
Number of answers: 423367

Rank	Target	Score	Precision	Count	Unfunny	Somewhat Funny	Funny
1	I'd better give it a little longer. It's a really tough case.	2.0921	0.0135	3107	751	1319	1037
2	He told me to cook the books, but didn't say for how long.	1.9946	0.0113	4458	1280	1922	1256
3	Hopefully cooking the books works better than laundering the money.	1.9777	0.0177	1880	577	768	535
4	Something t s					1568	1336
5	No, I won't be going back to work on Monday. Why	1.9679	0.0123	3864	1193	1602	1069

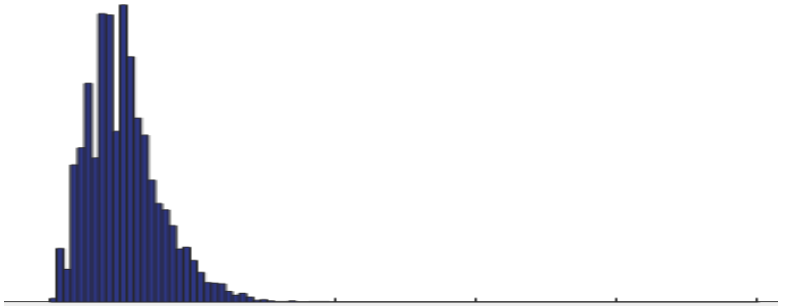
Data from contests:
https://github.com/nextml/caption-contest-data

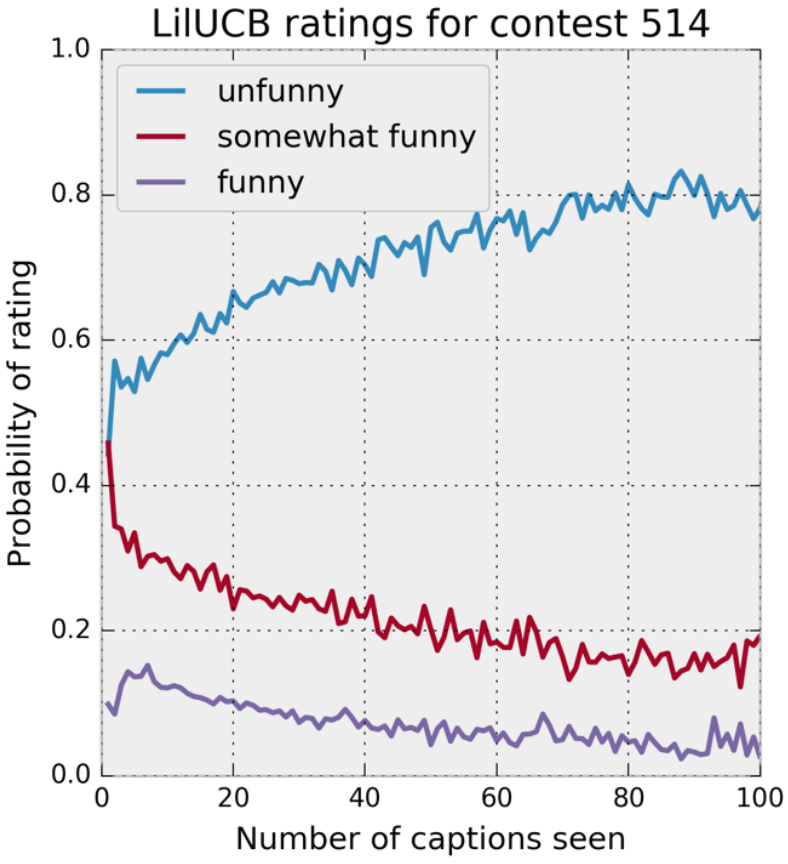
Benefits

4x times fewer ratings needed



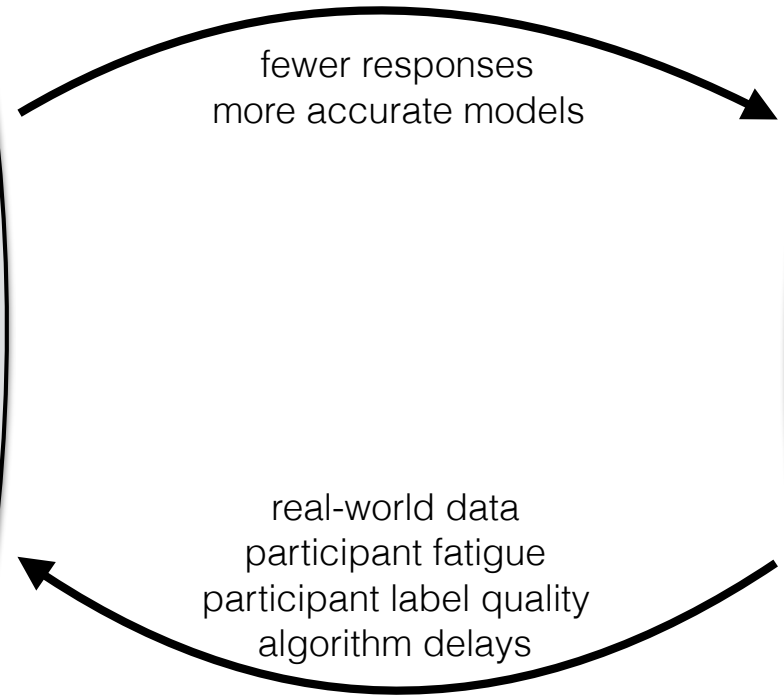
Data





Goal

Adaptive sampling algorithms



Crowdsourcing

Goal: let both parties *easily* use NEXT

Software uses

By default, NEXT can be applied to 3 problems

Select face on the bottom most similar to the face on top

Pool based triplets

Cardinal Bandits

comic by P. C. Vey

Dueling Bandits

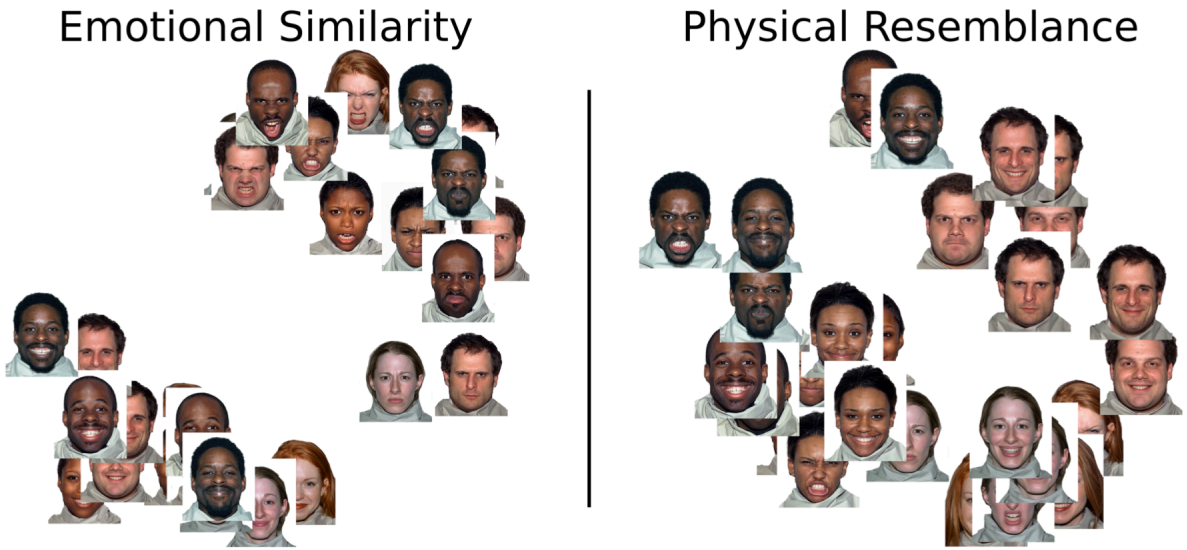
Select the street that looks safer

NEXT can also be used with REST API

Result requirements

1. Web browser
2. Amazon AWS account
3. ZIP of targets (e.g., images)
4. Experiment description (which has good documentation!)

After NEXT link sent to crowdsourcing service, results can be generated!



More detail on documentation:
https://github.com/nextml/NEXT/wiki

Key messages

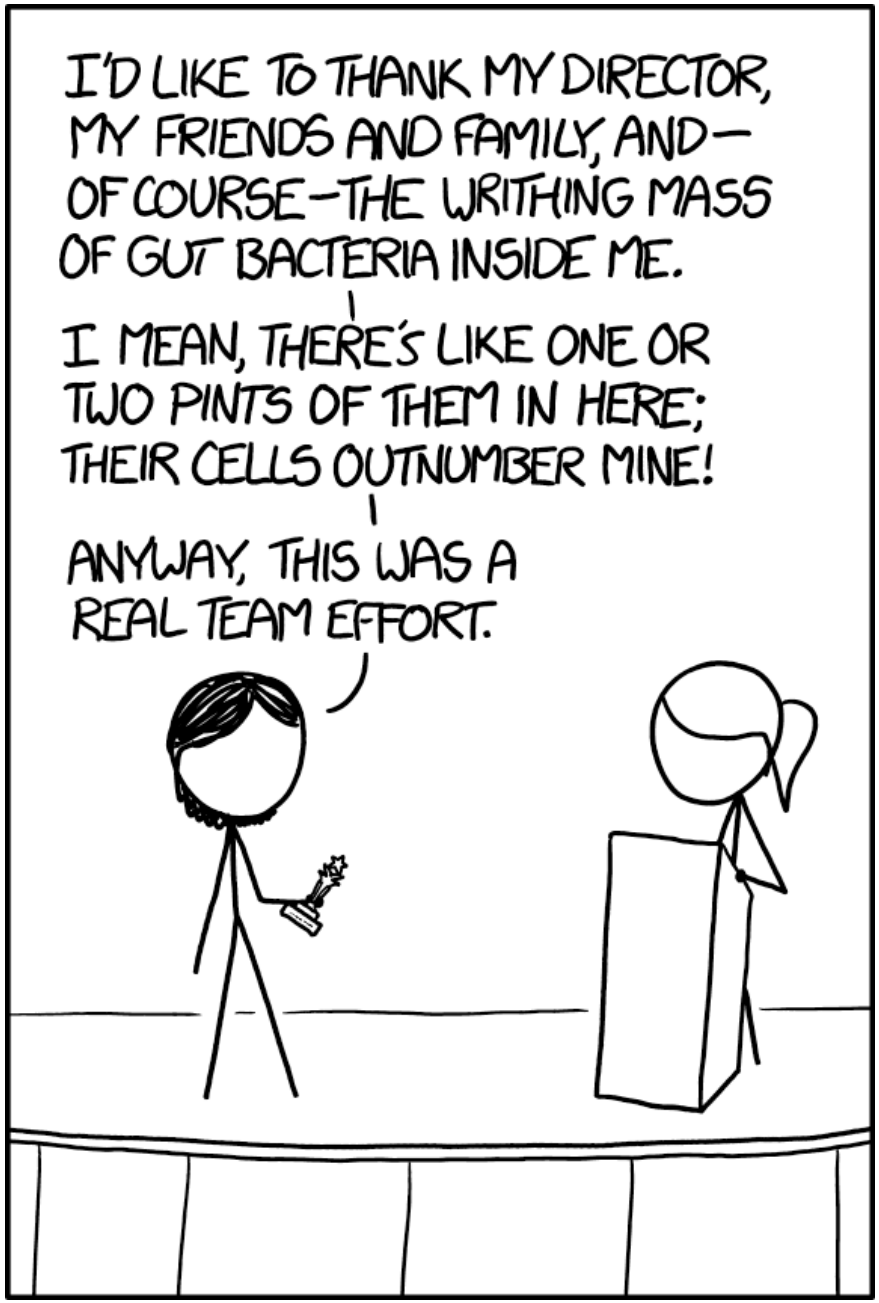
1. Adaptive sampling reduces data collection cost.

2. NEXT is a crowdsourcing data collection tool that can use adaptive sampling techniques

3. NEXT is easy* to use by experimentalists, algorithm developers and practitioners, and a mathematical background is not required.

4. NEXT developers experimentalist engagement to aid research and to gain feedback to improve the software

* NEXT has been created by an academic research group in collaboration with psychologists



https://xkcd.com/1543/