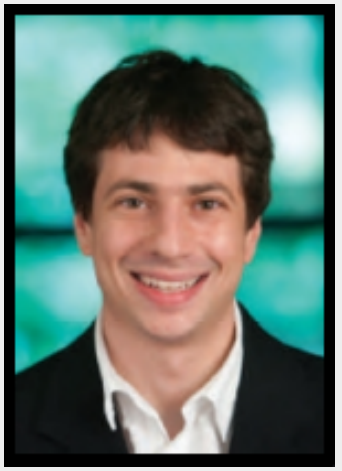
Local Differential Privacy For Evolving Data

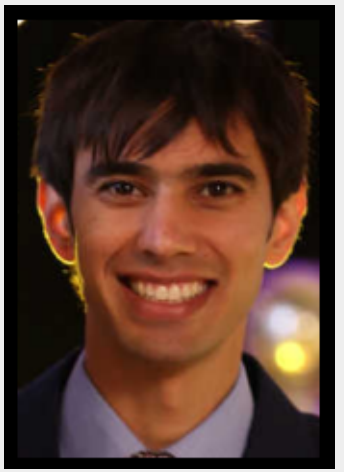
Matthew Joseph

Joint work with



Aaron Roth

Jonathan Ullman



Bo Waggoner

Penn

Northeastern University

University of Colorado Boulder

High-Level Motivation

Users have data and *analyst* wants to learn about data

But users want to keep their data private

How can analyst *learn from private data*?

What if private data is *evolving over time*?

What is “private”?

Here, *local differential privacy*

Each user randomizes own data

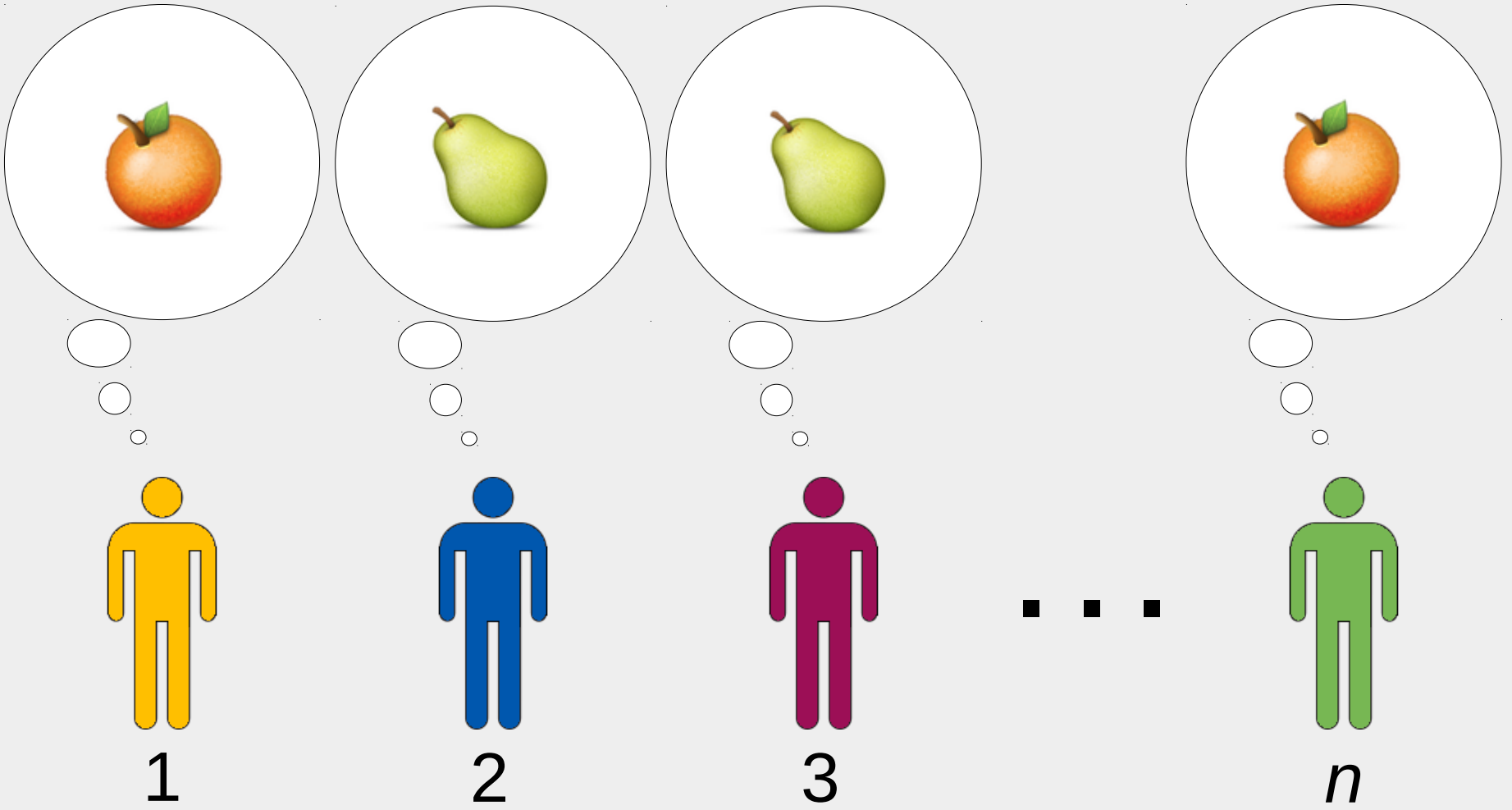
Analyst learns on randomized data

Randomness obscures individual data

Example

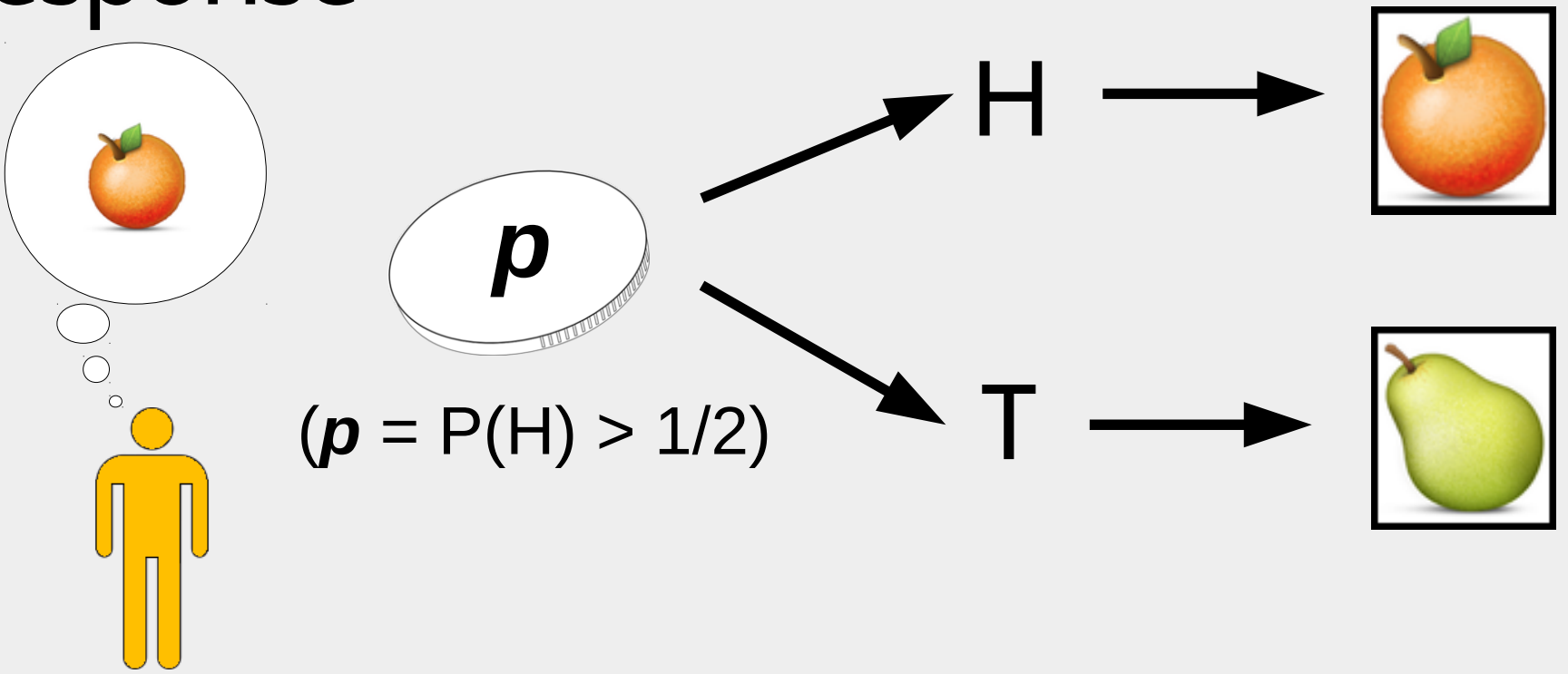
Do customers prefer oranges or pears?

Oranges or Pears?

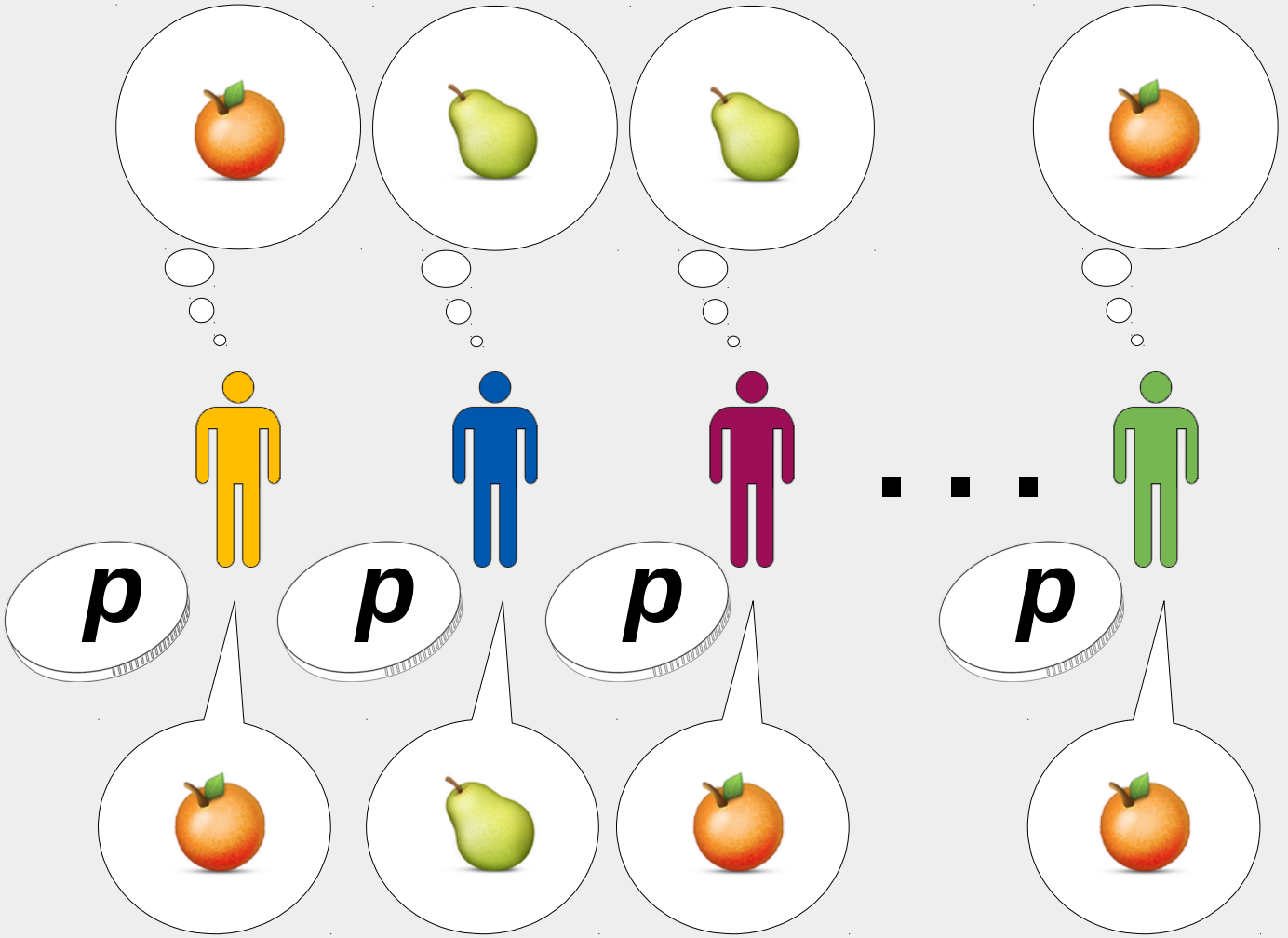


Oranges or Pears?

Known solution: randomized response



Oranges or Pears?



Oranges or Pears?

Add up responses

Most common response is answer

With high probability, answer is correct

Oranges or Pears?

Learn about *population*, not individual users

Users get privacy through *plausible deniability* of random coins

Higher ***p*** → more accurate, but less private

New Problem: *Evolving* Data

What if you want to track data changes over time ("evolving data")?

Old Solution vs. New Problem

Just do randomized response every day?

Problem: this may reveal private data over time (privacy loss "adds up")

Can the analyst stay up to date without compromising individual user data?

This Paper: New Solution For New Problem

Yes! Solution: users “vote” on when to update out-of-date statistics

Analyst can track distribution and guarantee that users only “lose privacy” for distribution changes

Worst-case privacy guarantee always holds

When data comes from appropriate (evolving) distribution, get a formal accuracy guarantee too

arxiv.org/abs/1802.07128 | Poster 153