CS286r: Topics at the Interface between Computer Science and Economics

Fall 2012

Information, Prediction, and Collective Intelligence

Yiling Chen

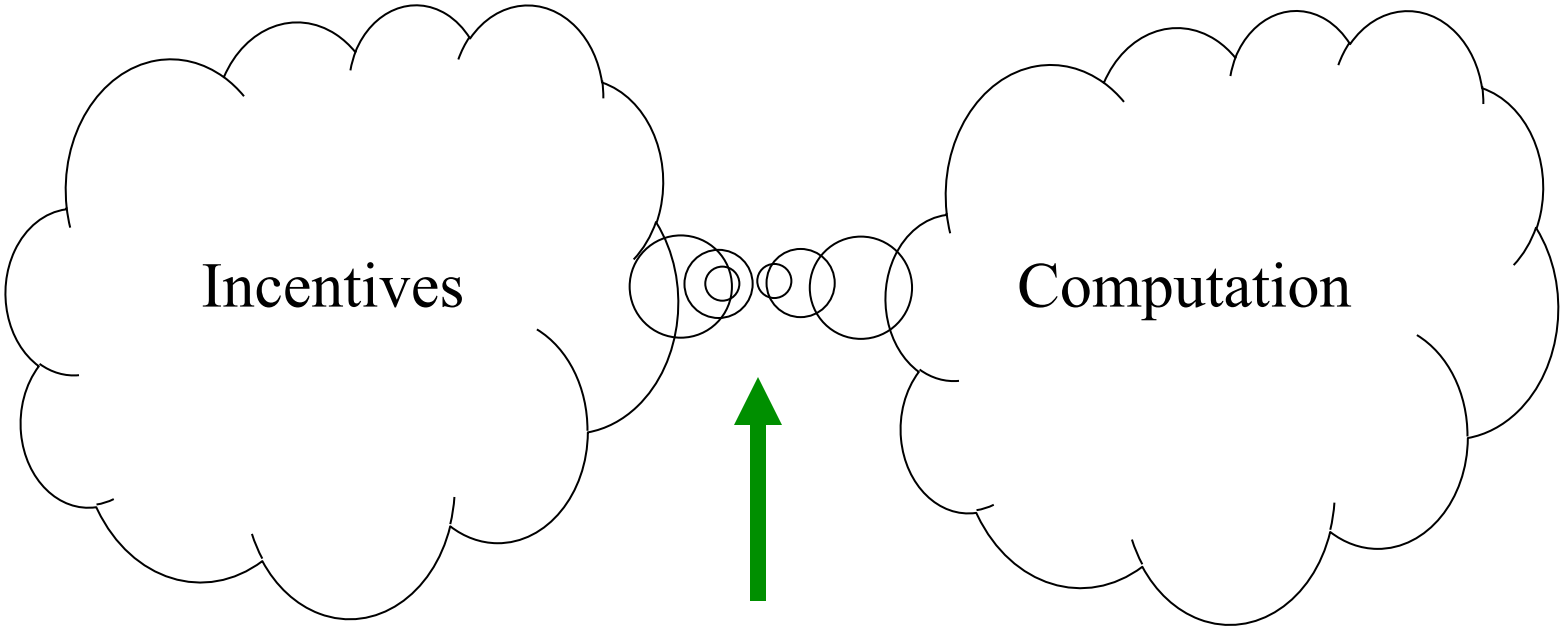
yiling@seas.harvard.edu

Course website: http://www.eecs.harvard.edu/cs286r/

Today's Plan

- CS 286r Fall 2011 topic and syllabus
- Example ideas and issues
- A bit background on decision making under uncertainty

Economics & Computation



Seek tractable interface

Theories, algorithms, and systems that satisfy both economic and computational constraints.

Lots of Compelling Applications

- Internet Monetization:
  Google, Yahoo!, Microsoft are using auctions to sell ads

- Markets are used for information aggregation
  - Google, Yahoo!, Microsoft, GE, etc. have internal prediction markets

- Social network:
  Facebook, Twitter, LinkedIn, Flickr, LibraryThing

- Peer-to-Peer systems
- Reputation systems

  …

This Course

- Rotating topic course
- Previous
  - Fall 2011. Computational Social Choice
  - Fall 2010. Information, Prediction, and Collective Intelligence
  - Fall 2009. Assignment, Matching, and Dynamics
  - Fall 2008. Social Computing
  - Spring 2008. Computational Finance
  - Spring 2007. Computational Mechanism Design
  - Spring 2006. Multi-agent Learning and Implementation

  …
- Seminar style

CS 186 in Spring is an introductory course to the area of economics and computation.

Course Goals

- Provide an introduction to an emerging, interdisciplinary literature
- Develop a level of comfort with both economic and computational thinking
- Develop general skills related to reading papers, identifying research questions
- Provide a basis for continued research.

Fall 2011

- Information, Prediction, and Collective Intelligence

- Algorithmic, game theoretic, and conceptual questions related to obtaining information, making predictions, and getting tasks done by the crowds.

Crowds Are Smarter…

- Who wants to be a millionaire?
  - Fifty-Fifty

    Correct 50% of the time
  - Phone-A-Friend

    Correct 65% of the time
  - Ask the Audience

    Correct 91% of the time

Crowds Are Smarter…

- Jelly-Beans-in-the-Jar Experiment
  - Professor Jack Treynor ran the experiment in his class
  - with a jar that held 850 beans
  - the group estimate was 871
  - only one of the 56 people in the class made a better guess

Are Crowds Smarter?

- No always
  - Bad committee decisions
  - Endless group meetings

- In this course, we focus on mechanisms that intend to make crowds smarter.

Structure of the Course

- Introductory lectures (6 lectures)
  - This one, game theory (2), and basics of proper scoring rules and prediction markets (3)
- Research Papers
  - Prediction Markets
  - Crowdsourcing and User-Generated Content
  - Peer Prediction
- One or two guest lectures

Prerequisites

- Math background is important! At least a basic course in linear algebra (such as M 21b, AM 21b, or equivalent)
- A course on probabilities and statistics (STAT 110 or equivalent)
- An algorithm course (CS 124, or equivalent)
- Familiarity with the concept of rationality. An AI course or an economics/game theory course.

  CS 186 and advanced course in algorithms, microeconomics, game theory are helpful but **not** required.

Reading Materials

- For each class, we have provided some reading materials. We ask you to read them and submit your comments by midnight before class.

- There will be reading questions for each class.
  - Your comments should include good faith answers to these questions.
  - The questions are designed to facilitate in understanding or to encourage discussion.

Grading

Problem sets	25%	2 homework problem sets
Participation	25%	Reading papers, submitting reading comments and questions before class, and participation in class discussion. (Note: Absent students rarely contribute to discussions.)
Presentation of one or two sets of research papers	15%	A short survey and critique of the papers. See presentation notes. Lead class discussion.
Project	35%	Project proposal, class presentation, and final report.

Project

- **Goal:** develop a deep understanding of a specific research area and to the extend possible to work on an open research problem.
- Can be theoretical, computational, experimental, or empirical.
- Can write an exposition paper, but needs novelty!
- Tentative project due dates:
  - Tuesday 10/30: project proposal due
  - Wednesday 12/5: brief project presentation
  - Friday 12/7: project report due

Logistics

- TF
  - Mike Ruberry

- Office Hours
  - Yiling: Monday 2:30 – 3:30, MD 339
  - Later will add office hours likely on Thursdays to meet with students in advance of presenting papers
  - Mike: Wed 11---12, MD second floor lounge

Missed course materials from the TF

Example Ideas and Issues

Events of Interest

- Will category 3 (or higher) hurricane make landfall in Florida in 2011?
- Will Google reinstate its Chinese search engine?
- Will Democratic party win the Presidential election?
- Will Microsoft stock price exceed $30?
- Will there be a cure for cancer by 2015?
- Will sales revenue exceed $200k in April?

……

Incentivize Experts

- Suppose I'd like to get information about tomorrow's weather (sunny or rainy?)

- How can I ensure that an expert will tell me his/her **true** probability assessment of the event?

Proper Scoring Rules

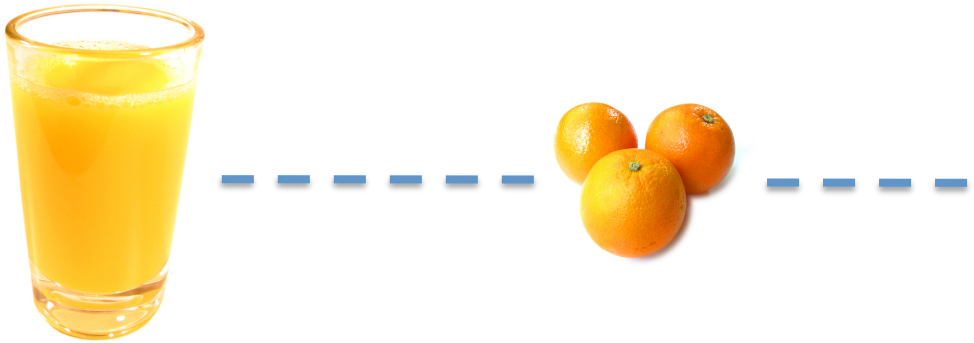
Combining information is hard!

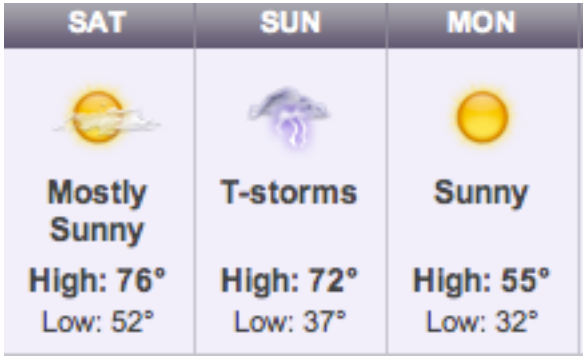
- If we have multiple experts, how can we combine their information?
- Some impossibility results on combining probability distributions.
  - T(f1, f2, f3, …, fn)
  - External Bayesianity
  - Independent of irrelevant alternatives

  => dictatorship

Orange Juice Futures and Weather

Trades of 15,000 pounds of orange juice solid in March





- Orange juice futures price can improve weather forecast! [Roll 1984]

Bet = Credible Opinion

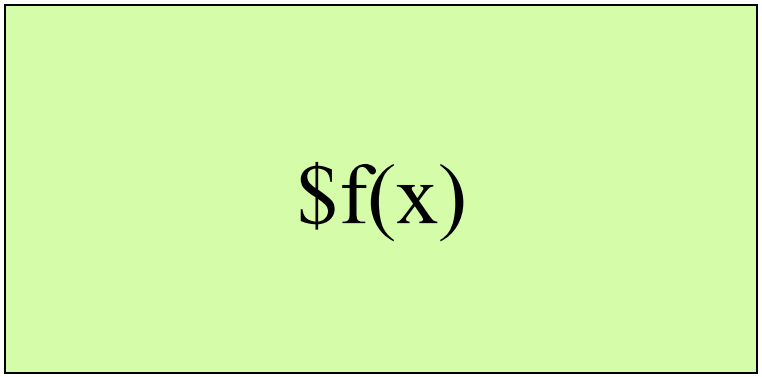
- Q: Is Vinay Deolalikar's proof of P≠NP correct?

  "If Vinay Deolalikar is awarded the $1,000,000 Clay Millennium Prize for his proof of P≠NP, then I, Scott Aaronson, will personally supplement his prize by the amount of $200,000."

- Scott Aaronson: "I have a way of stating my prediction that no reasonable person could hold against me: I've literally bet my house on it."

Prediction Markets

- A prediction market is a futures market (betting intermediary) that is designed for information aggregation and prediction.
- Payoffs of the traded item is associated with outcomes of future events.



Intrade

Barack Obama to be re-elected President in 2012

Last prediction was: **$5.87** / share

Today's Change: ▲ **+$0.02** (+0.3%)

Contract Type: **0-100**

58.7%

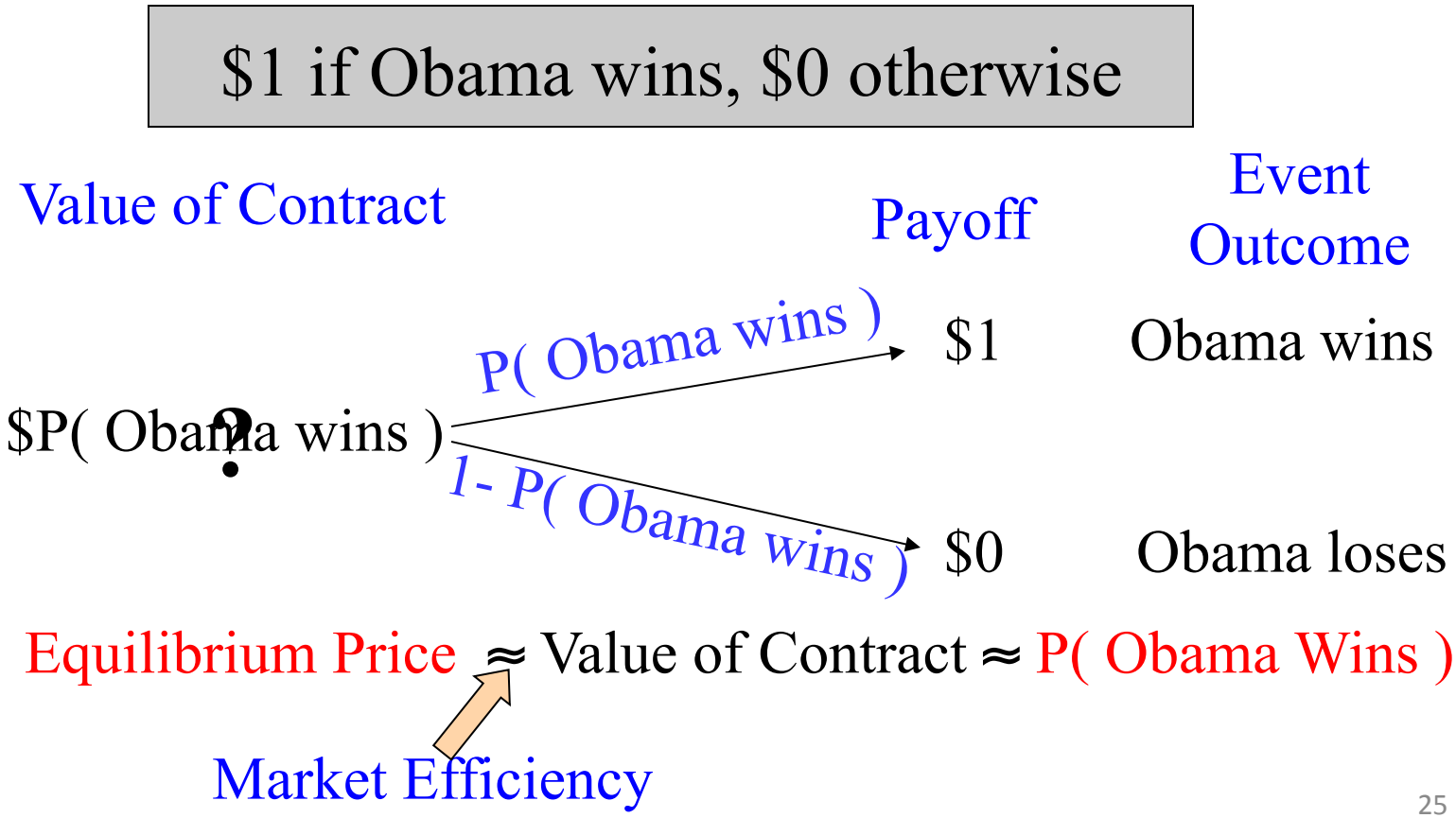
CHANCE

Event: 2012 Presidential Election Winner (Individual)

Function of Markets 1: Get Information

- Speculation ➔ price discovery

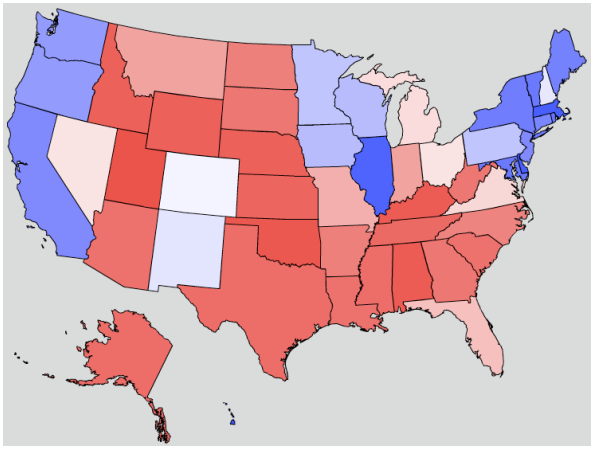
  price ≈ expectation of r.v. | all information



A Combinatorial Betting Example

- 2^{51} outcomes, $2^{2^{51}}$ combinations
- Allow participants to bet on logical formulas
  - Create contracts on the fly:

    $1 if Ohio AND Florida OR New York, $0 otherwise
  - Specify buy price and quantity
- Computationally hard!



We will look at

- Design (better) market mechanisms for information aggregation (connecting to proper scoring rules)
- Characterizing information aggregation with rational agents
- Enabling combinatorial markets

What If We Won't Know the Outcome?

- Eg. Conditional events, subjective information
- Surveys
  - Eg. How many hours per week you spent on assignments?
    - Less than 5 hours
    - 5-10 hours
    - 10-20 hours
    - Above 20 hours

Peer Prediction and Bayesian Truth Serum

Organized Human Computation

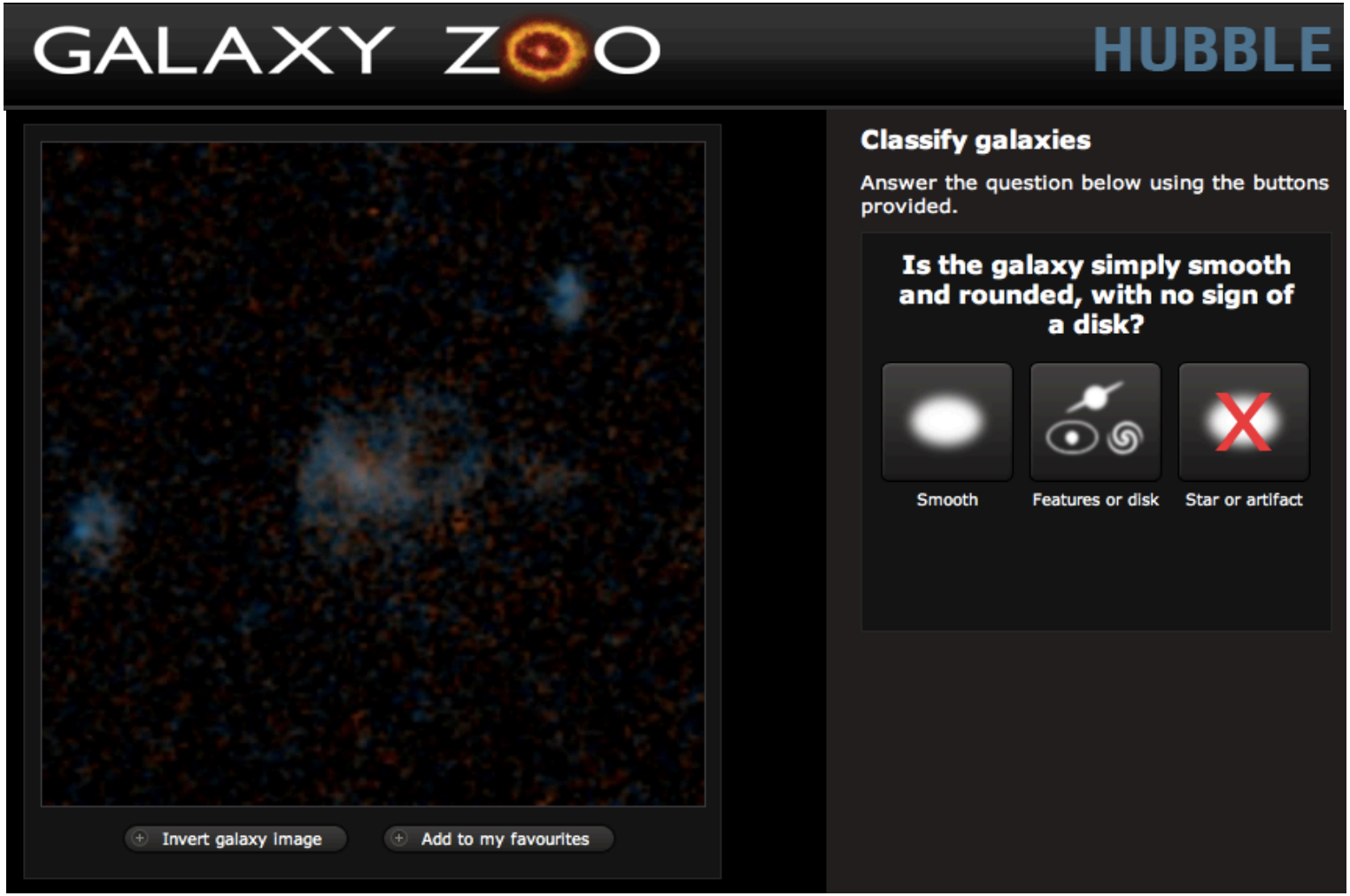
- An old idea
  - Halley's Comet (1758)
    - 3 astronomers calculate the trajectory of the Halley's Comet
  - The Math Table Project (1938-1948)
    - 450 out-of-work clerks
- Computer: a person who performs calculation as a profession
- Given a "computing plan"
- Quality assurance: computation was done by two independent human computers and checked by a third

The web changes everything …

The ESP Game

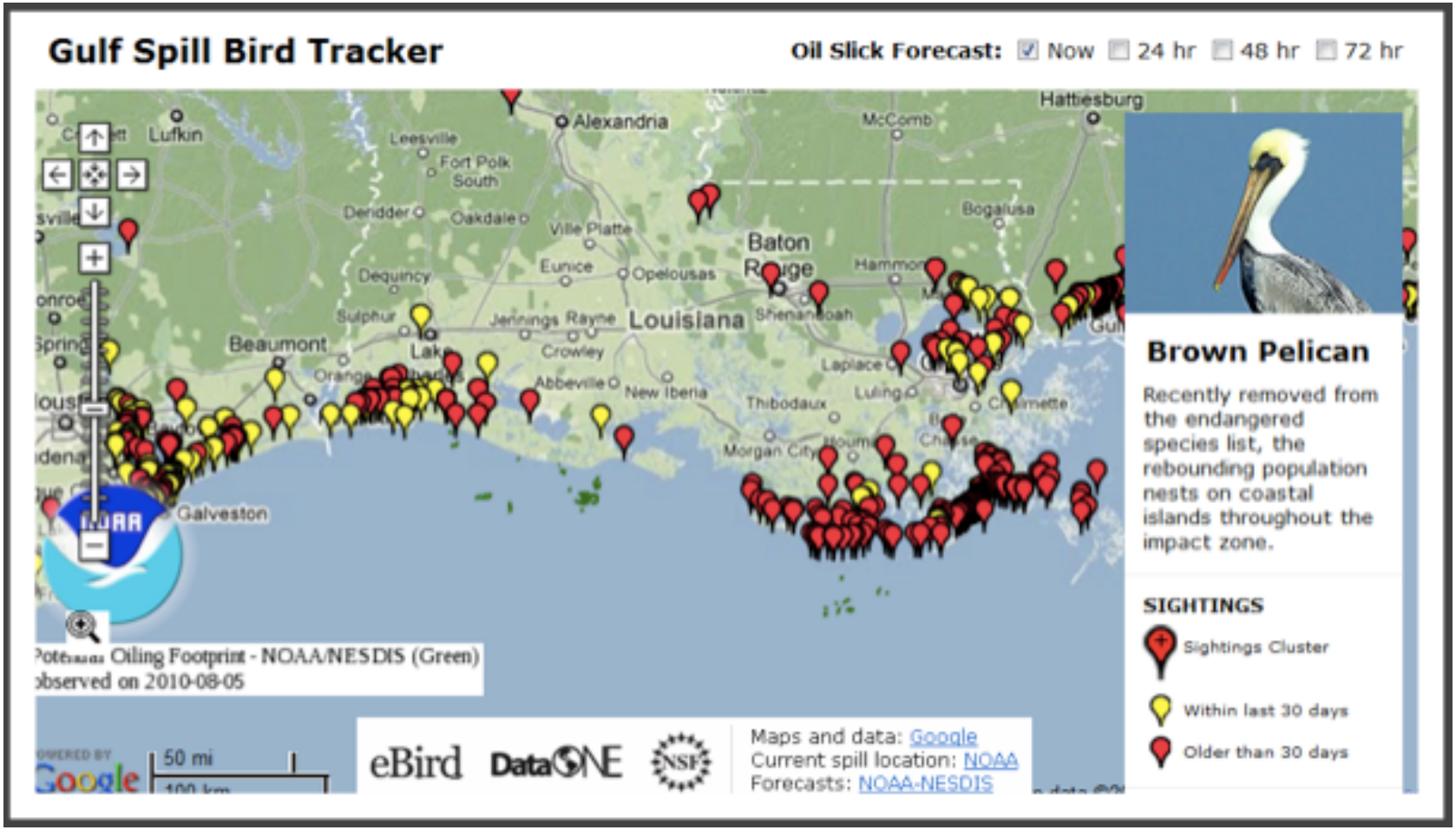

score
200
ESP Game
Concentrate...
time
2:25
What do you see?
taboo words
bison
dozer
guesses
kid
submit
pass

Galaxy Zoo



More than 200,000 participants from 113 countries; more than 100 million classifications

eBird


Gulf Spill Bird Tracker
Oil Slick Forecast: Now 24 hr 48 hr 72 hr
Alexandria
Lufkin
Leesville
Fort Polk South
Deridder
Oakdale
Ville Platte
Eunice
Opelousas
Dequincy
Sulphur
Jennings
Rayne
Louisiana
Crowley
Beaumont
Lake
Abbeville
New Iberia
Baton Rouge
Hammond
Bogalusa
McComb
Hattiesburg
Thibodaux
Morgan City
Laplace
Luling
Galveston
LAURA
Oiling Footprint - NOAA/NESDIS (Green)
observed on 2010-08-05
50 mi
Google
eBird
DataONE
NSF
Maps and data: Google
Current spill location: NOAA
Forecasts: NOAA-NESDIS
Brown Pelican
Recently removed from the endangered species list, the rebounding population nests on coastal islands throughout the impact zone.
SIGHTINGS
Sightings Cluster
Within last 30 days
Older than 30 days

Amazon Mechanical Turk (Mturk)

All HITs

1-10 of 2372 Results

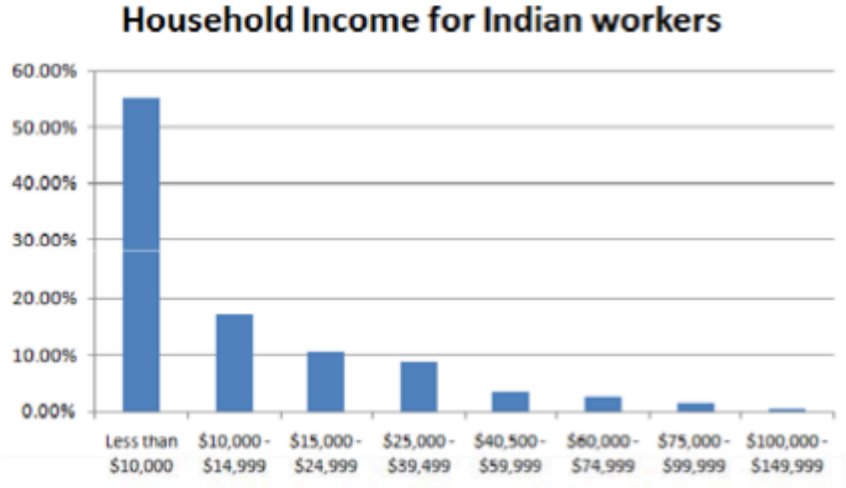
Sort by: HITs Available (most first) GO!

Show all details | Hide all details

1 2 3 4 5 › Next » Last

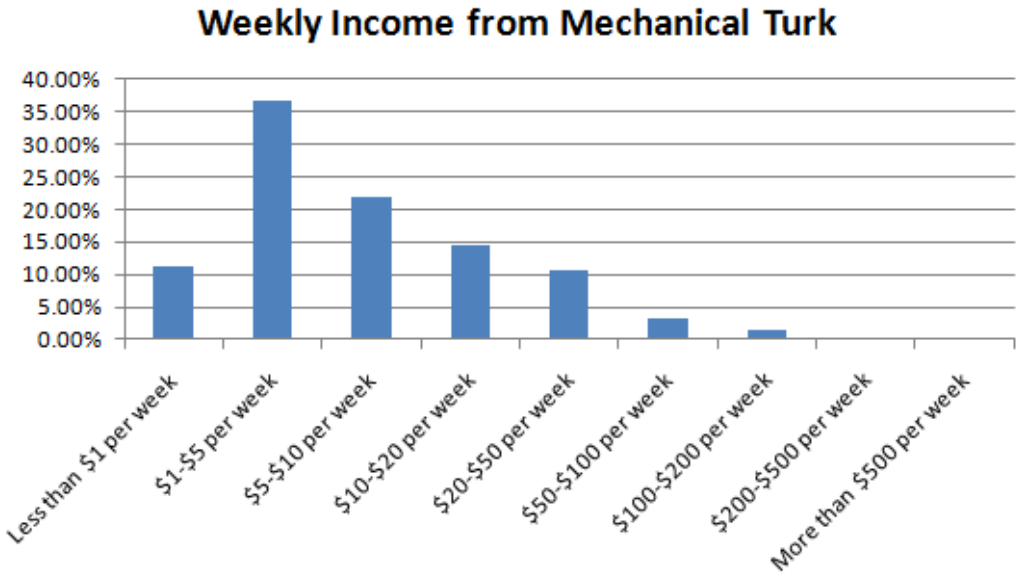
Give Your Opinion - Simple and Quick! (US)				View a HIT in this group
Requester: CrowdSource	**HIT Expiration Date:**	Jun 27, 2013 (52 weeks)	**Reward:**	$0.16
	Time Allotted:	32 minutes	**HITs Available:**	14906
find email for given person via web search (~60sec) approved regularly			Not Qualified to work on this HIT (Why?)	View a HIT in this group
Requester: Sebastian Darr	**HIT Expiration Date:**	Jul 30, 2012 (4 weeks 4 days)	**Reward:**	$0.03
	Time Allotted:	5 minutes	**HITs Available:**	8488
Get researcher's email address (~60sec) daily approval!			Not Qualified to work on this HIT (Why?)	View a HIT in this group
Requester: Sebastian Darr	**HIT Expiration Date:**	Jul 27, 2012 (4 weeks 1 day)	**Reward:**	$0.03
	Time Allotted:	5 minutes	**HITs Available:**	8180
web search for email address (~60sec) daily approval!			Not Qualified to work on this HIT (Why?)	View a HIT in this group
Requester: Sebastian Darr	**HIT Expiration Date:**	Jul 27, 2012 (4 weeks 1 day)	**Reward:**	$0.03
	Time Allotted:	5 minutes	**HITs Available:**	6684

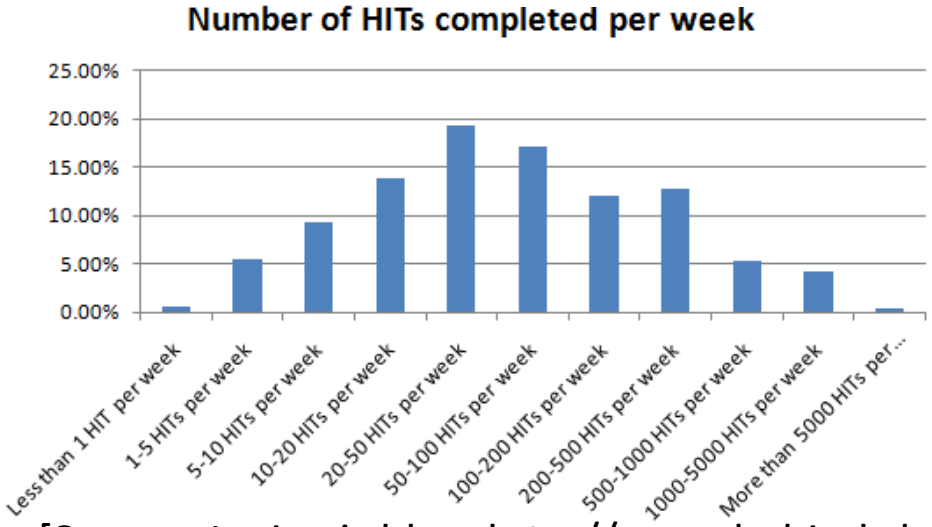
Demographics of Turkers

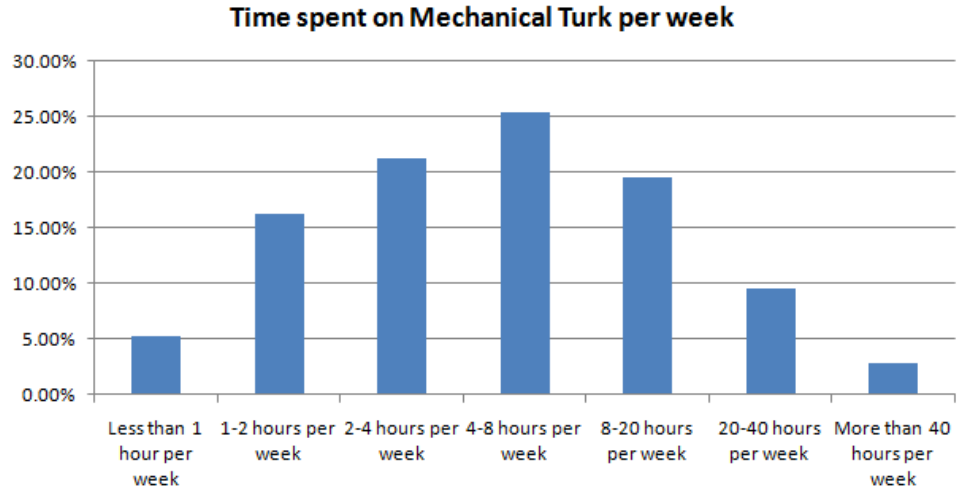


[Source: Ipeirotis blog, http://www.behind-the-enemy-lines.com/2010/03/new-demographics-of-mechanical-turk.html]

Demographics of Turkers







[Source: Ipeirotis blog, http://www.behind-the-enemy-lines.com/2010/03/new-demographics-of-mechanical-turk.html]

We'll look at

- Quality and workflow control for crowdsourcing

- How to incentivize "better" contributions

A Bit Background on Decision Making under Uncertainty

Uncertainty, Risk, & Information

- Uncertainty

   or 

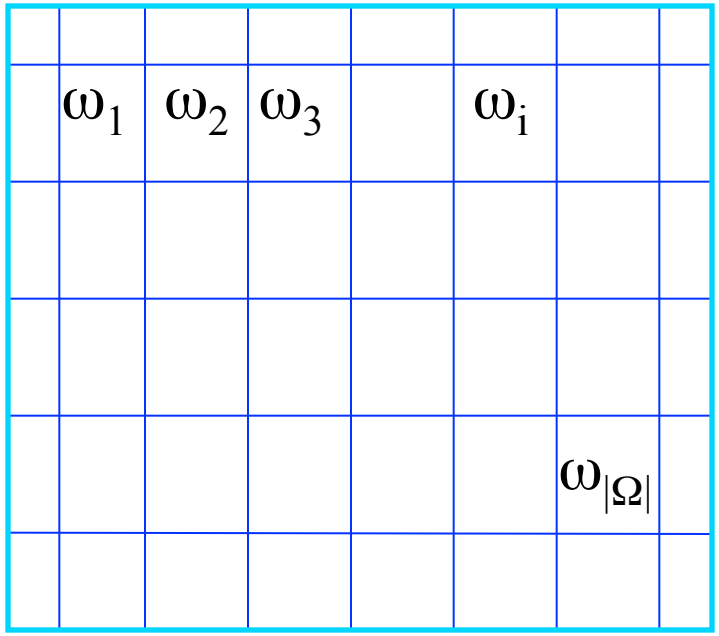
- Risk

  Pr() Pr()

- Information

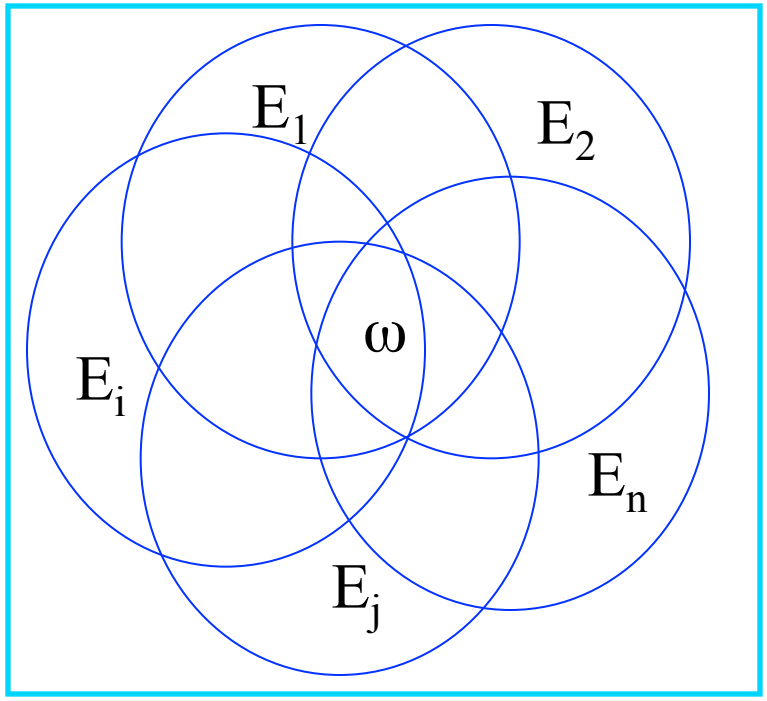
  Pr( | info) Pr( | info)

Uncertainty & Risk, in General



- Ω: State Space
- ω are disjoint exhaustive
  states of the world
- ω_j: rain tomorrow & have umbrella & ...
- Pr(ω)

Uncertainty & Risk, in General



Alternatively,

- Overlapping events
  - E1: rain tomorrow
  - E2: have umbrella

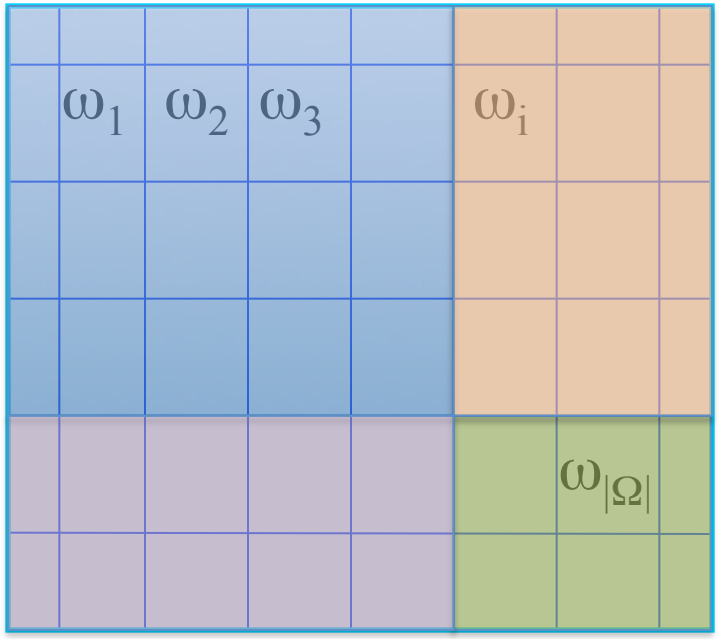
- $|\Omega|=2^n$

Modeling Information

- E: Event of interest
- $P(E, S_i, S_j)$: Prior distribution
- Nature draws event outcome and signals
- Bayesian agents can form belief

$$P(E=e|S_i = s_i)$$

An Economist's Approach to Modeling Information

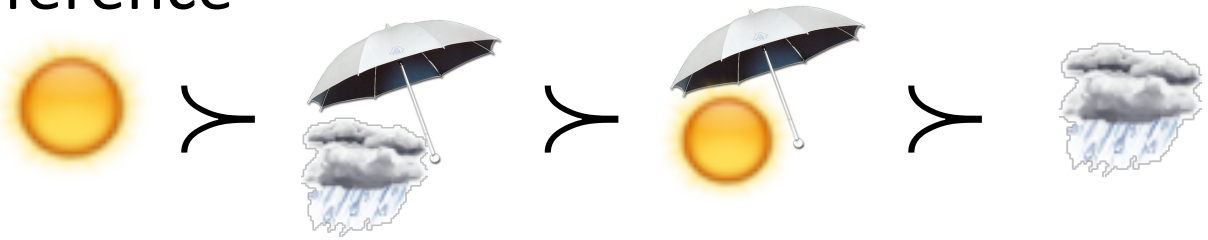


- Ω: state space
- $Pr(\omega)$
- An agent has a partition of the state space*
- Nature draws ω^*
- Agent observes $S_i(\omega^*)$
- Agent forms belief

$$P(\omega | S_i(\omega^*))$$

Preference and Utility

- Preference



- Utility, $u(\omega)$

u( )=10 >

u( )=8 >

u( )=-4 >

u( )=-10

Decision Making Under Uncertainty

- Maximize expected utility
  - $E[u]= \sum_{\omega} Pr(\omega)u(\omega)$
- Decisions (actions) can affect $Pr(\omega)$ or $u(\omega)$

					E[u]
Don't Take umbrella	0.5	0	0	0.5	.5*10+.5*(-10) =0
Take umbrella (but I may leave it at the library)	0.25	0.25	0.25	0.25	.25*10+.25*8+.25*(-4)+.25*(-10) =1

Should take umbrella!

Utility of Money and Risk Attitude

- Outcomes are $

- Risk attitude:
  - risk neutral: $u(x) \sim x$
  - risk averse (typical):
    u concave ($u''(x) < 0$ for all x), e.g. $u(x)=log(x)$
  - risk prone: u convex

Risk Attitude & Hedging

- I'm risk averse, u(x) = log (x), insurance company A is risk neutral, u(x)=x.
- I believe that my car might be stolen with prob. 0.01

ω_1: car stolen	ω_2: car not stolen
$u(\omega_1) = \log(10{,}000)$	$u(\omega_2) = \log(20{,}000)$

E[u]=.01 (4)+.99 (4.3) = 4.2980

- I buy $10,000 insurance for $125

$u(\omega_1) = \log(19{,}875)$	$u(\omega_2) = \log(19{,}875)$

E[u]=.01 (4.2983)+.99 (4.2983) = 4.2983

- Insurance company A also believes Pr(car stolen)=0.01

$u(\omega_1) = -9{,}875$	$u(\omega_2) = 125$

E[u]=.01 (-9875)+.99 (125) = 25 >0

I am happy to buy insurance. Insurance company A is happy to sell it. The transaction allocates risk.

Probability and Speculating

- Suppose that I'm also risk neutral, u(x)=x.
- But I think that the probability for my car being stolen is much higher than 0.01, say 0.1.
- A $10,000 car insurance is worth

  .1 (10,000)+.9 (0)=$1,000

  to me, but the insurance company only asks for $125.

  Too cheap!
- Buy the insurance, and I get $825 on expectation.

I am speculating the insurance company.

For Mon. 9/10

- Lecture on games with complete information
- Submit comments on Chapter 3 and Chapter 5.1 of Multiagent Systems book before midnight 9/9
  - download the readings from the Schedule page of the course website
  - Those who took 286r in Fall'11, readings are different and classes are optional for 9/10 and 9/12
- Reading questions will be posted on the Schedule page
- Please give TF your email address so that we can register you for the comment submission system