Harvesting Social Knowledge

Bernardo A. Huberman

Director, Systems Research

HP Labs

tapping the information economy

people are not just another asset

they are at times irrational, often strategic and different from each other

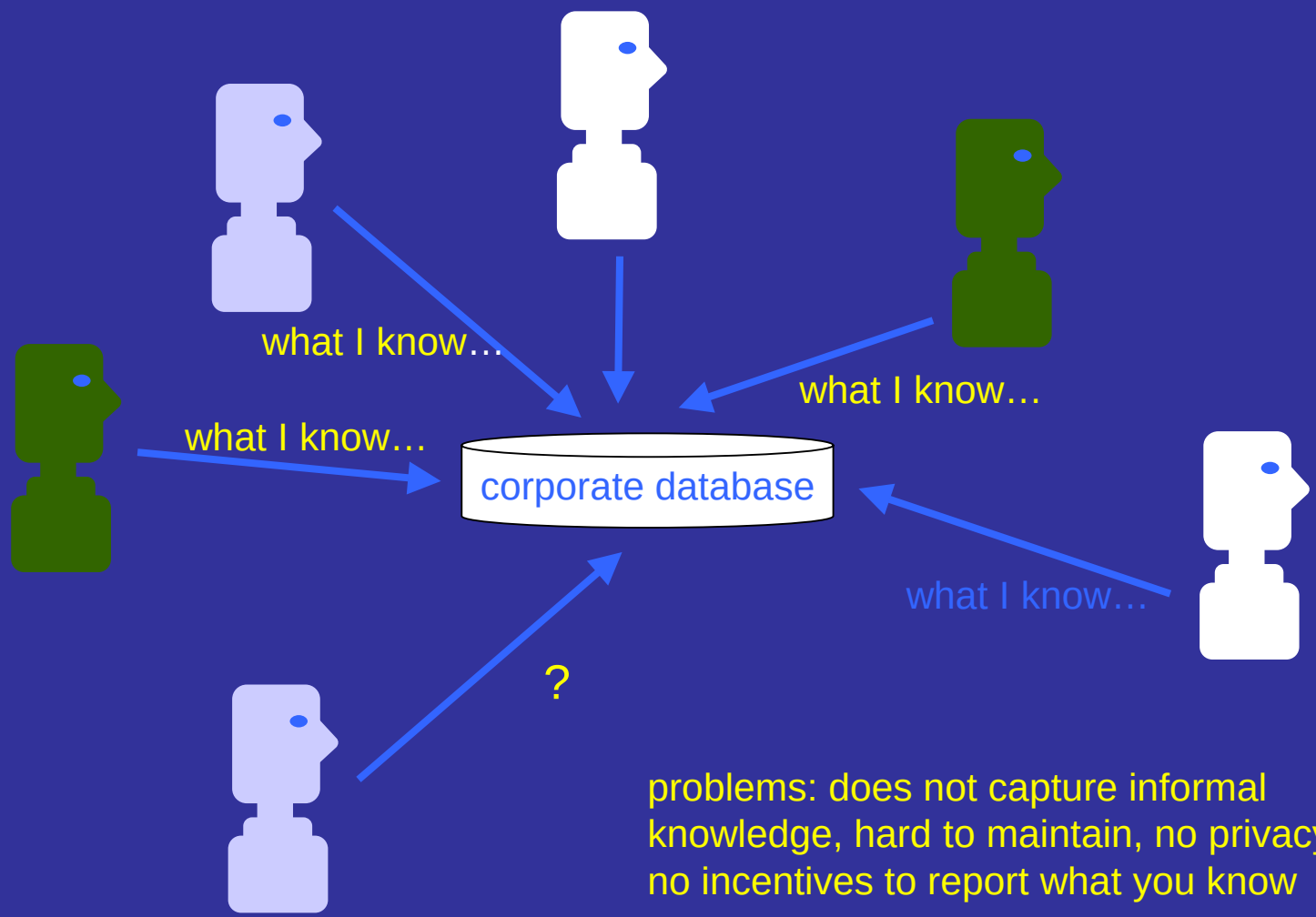
and they value their privacy

problems:

1. to find those we need, and to induce them to reveal information that is truthful.
2. to aggregate it in interesting ways
3. to respect their privacy

organizational knowledge

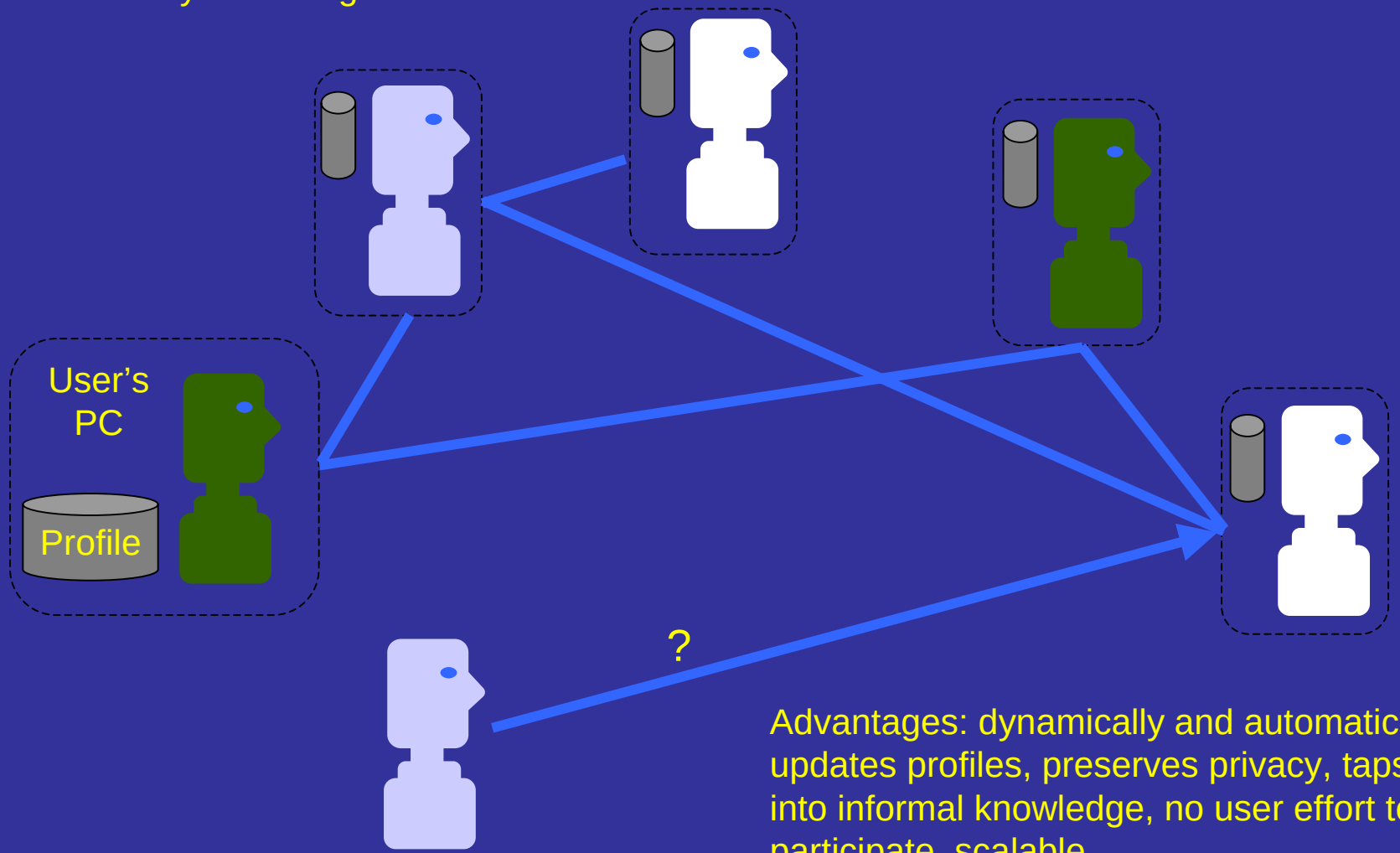
traditional solution



problems: does not capture informal knowledge, hard to maintain, no privacy, no incentives to report what you know

a novel way: p2p

shock: social harvesting of community knowledge



Advantages: dynamically and automatically updates profiles, preserves privacy, taps into informal knowledge, no user effort to participate, scalable

the future

we all care about it.

and invest resources in finding out about it.

Caravaggio ,The Fortune Teller, 1596-97

“it is hard to predict anything, especially the future”

Niels Bohr

how do organizations predict?

- they ask the experts (and consultants)
- have meetings (lots of them)
- designate someone as forecaster
- take a vote (not very good)

an alternative

- to tap into organizational knowledge
- distributed among individuals within the organization
- who often don’t realize that what they know is valuable
- how do we find them?
- and turn their knowledge into useful predictions?

markets

- markets aggregate and reveal information (Hayek, Lucas, etc.)
- to predict outcomes, use markets where the asset is information (rather than a physical good)
- examples
  - iowa electronic markets
  - hollywood stock exchange

markets within organizations

-problematic-

- low participation
- illiquidity
- information traps
- hard to motivate
- easily manipulated

still…

- some participants in information markets have superior knowledge, or are better processors, of the knowledge harnessed by the market
- also, individuals differ in their risk attitudes
- how do we find those behavioral attitudes?
- and take them into account when aggregating their predictions?

a new mechanism

(with kay-yut chen and leslie fine)

- it identifies participants that have these talents, and extracts their risk attitudes
- it induces them to be truthful
- while avoiding the pitfalls of small groups
- it aggregates information in nonlinear fashion

Proceedings of the 3rd ACM conference on e-commerce, 58-64 (2001)

two stages

stage 1: a market for contingent securities.

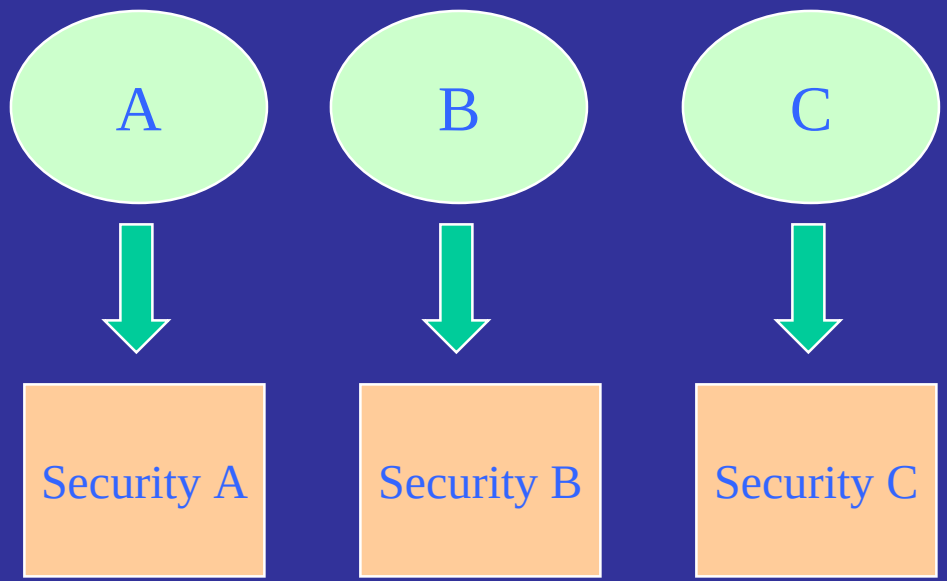
- it provides behavioral information, such as risk attitudes –synchronous-

stage 2: participants generate predictions on outcomes, which are then aggregated.

- incorporates behavioral information
- -asynchronous-

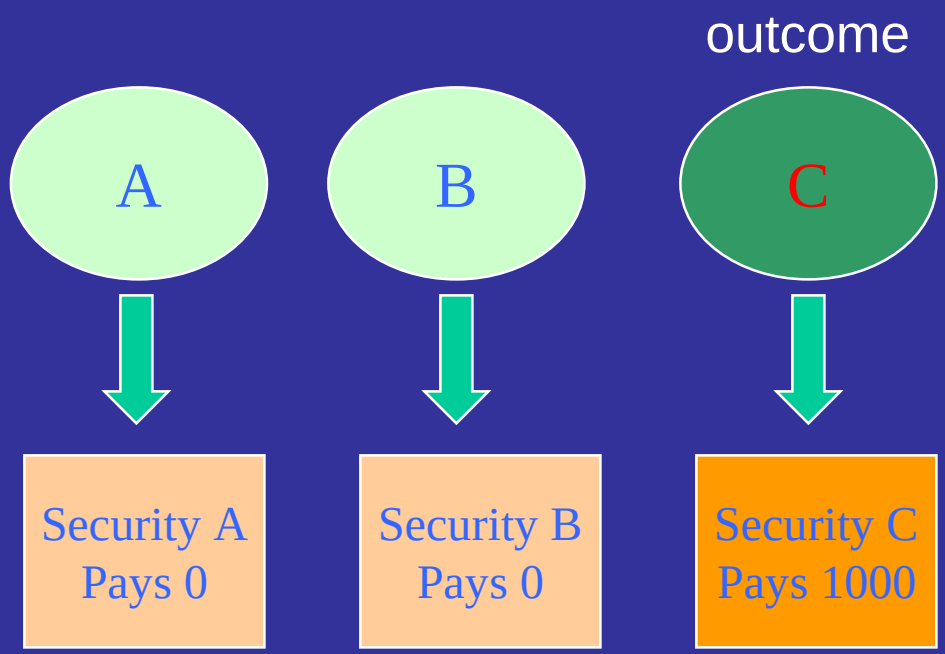
stage 1: an information market

three possible outcomes

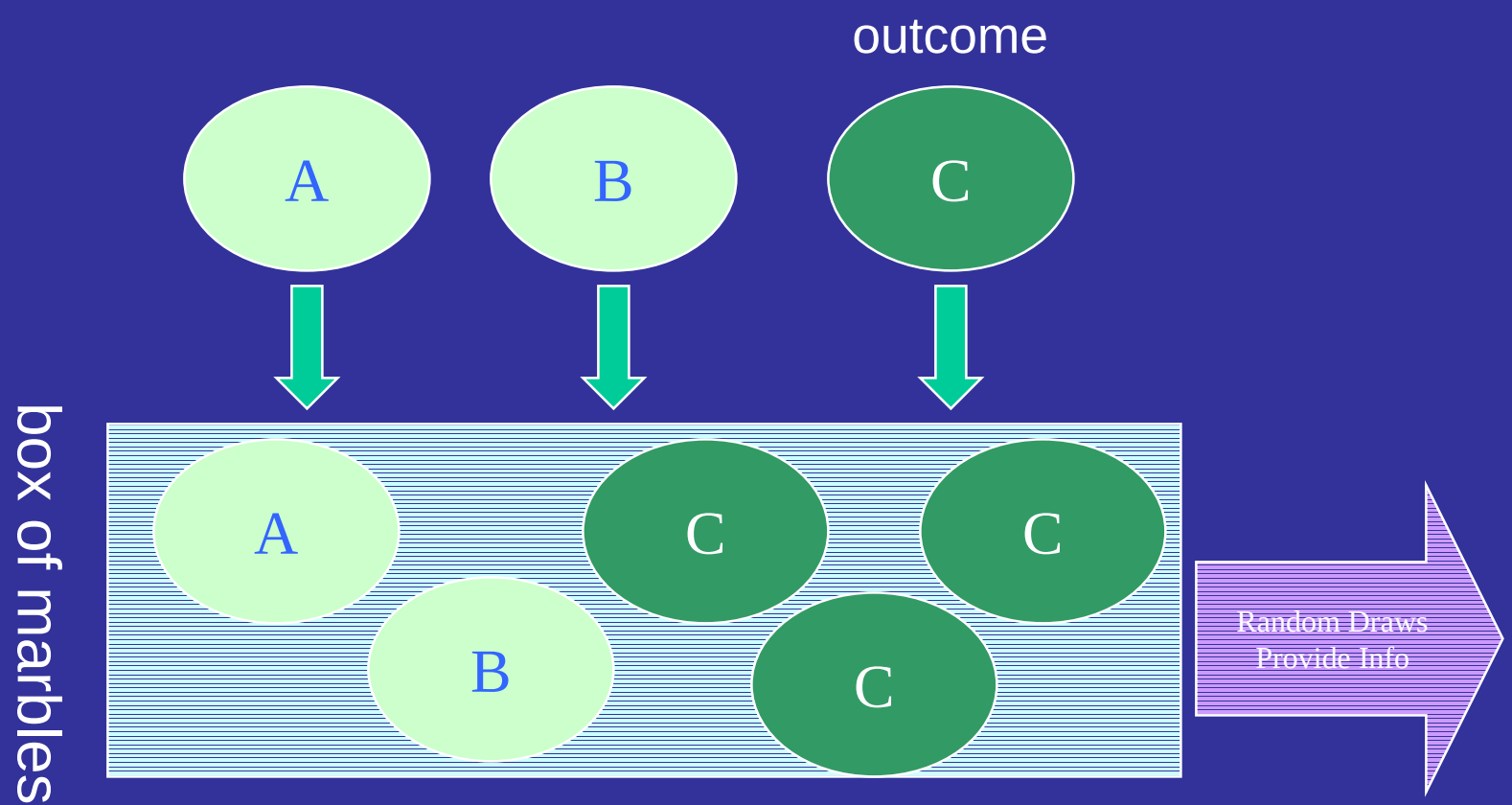


arrow-debreu

lottery-like securities



inducing diverse information



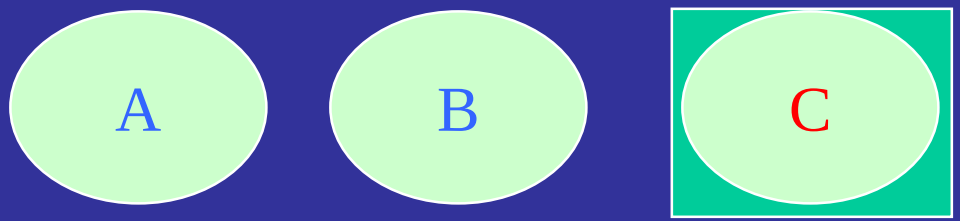
*** in actual experiments, there are TEN states**

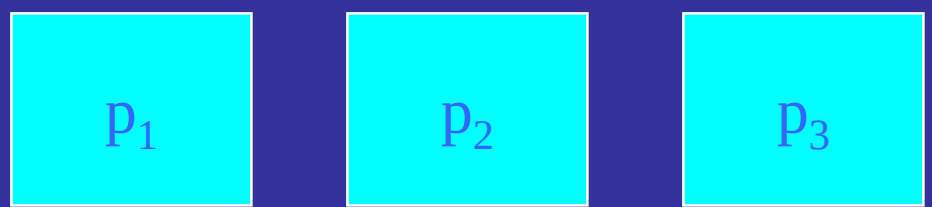
stage 2- forecasting

- participants are given 100 tickets
- to be allocated among 10 securities
- this determines probabilities
- true state pays according to the number of tickets allocated to it

inducing truth revelation

three possible outcomes





with nonlinear payoff

$\$ = C_1 + C_2 * \log(p_3)$

aggregating predictions

the probability of event *S* occurring, conditioned on *I*, is given by

$$P(s \mid I) = \frac{p_{s_1}^{\beta_1} p_{s_2}^{\beta_2} \cdots p_{s_N}^{\beta_N}}{\sum_{\forall s} p_{s_1}^{\beta_1} p_{s_2}^{\beta_2} \cdots p_{s_N}^{\beta_N}}$$

with β an exponent that denotes behavioral attitudes

>1 risk averse

<1 risk seeking

=1 risk neutral

what determines the exponent?

$$P(s \mid I) = \frac{p_{s_1}^{\beta_1} p_{s_2}^{\beta_2} \cdots p_{s_N}^{\beta_N}}{\sum_{\forall s} p_{s_1}^{\beta_1} p_{s_2}^{\beta_2} \cdots p_{s_N}^{\beta_N}}$$

normalization constant

$$\beta_i = r(V_i / \sigma_i)c$$

~sum of prices/winning payoff
It measures market risk

holding value/risk
- measures relative risk of individuals

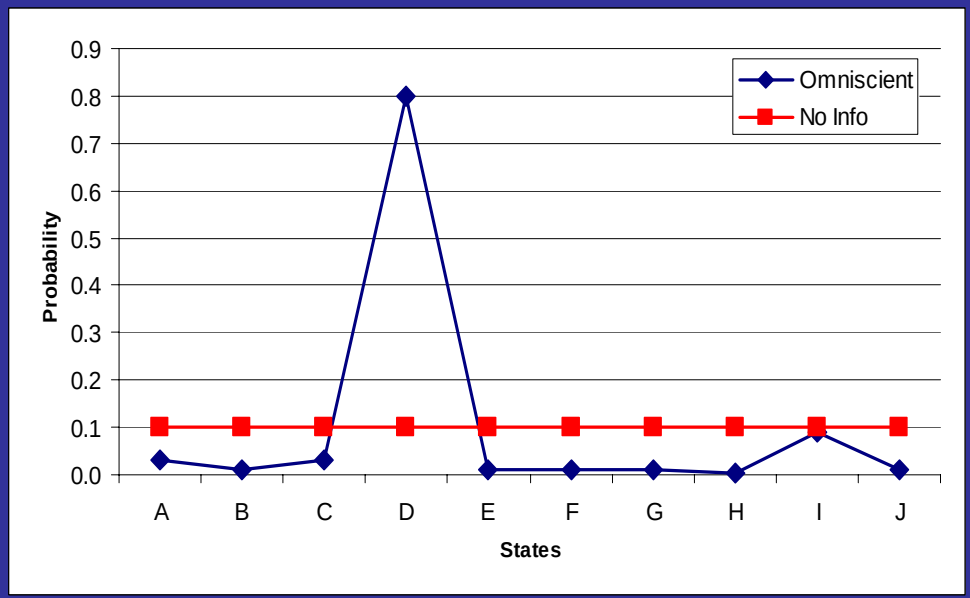
experiments

- human subjects in the laboratory (HP labs)
- each group receives diverse information
- run the two-stage mechanism
- and measure its performance

results

comparison to omniscient probability

Kullback-Leibler = 1.453

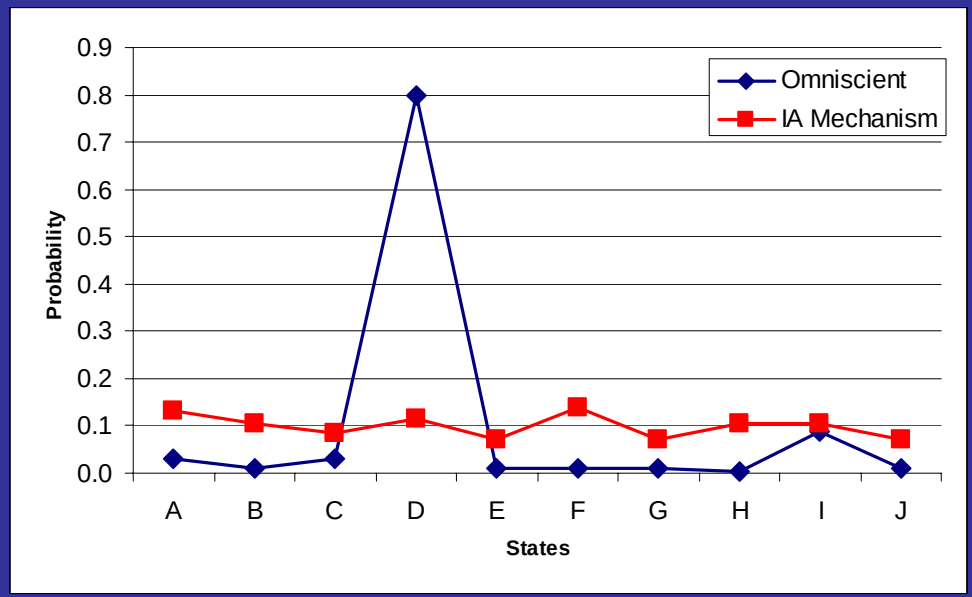


Experiment 4, Period 17
No Information

results

comparison to omniscient probability

Kullback-Leibler = 1.337

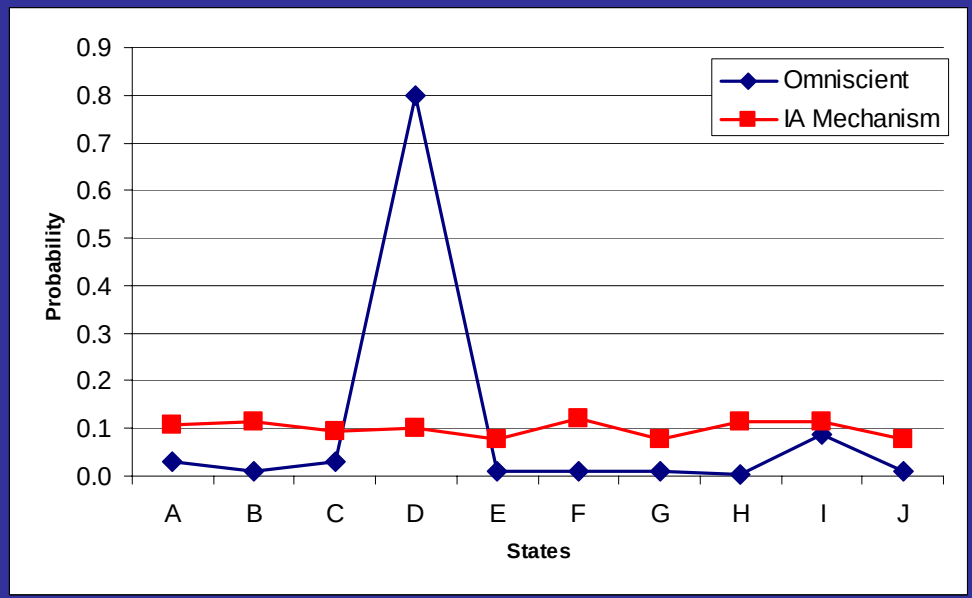


Experiment 4, Period 17
1 Player

results

comparison to omniscient probability

Kullback-Leibler = 1.448

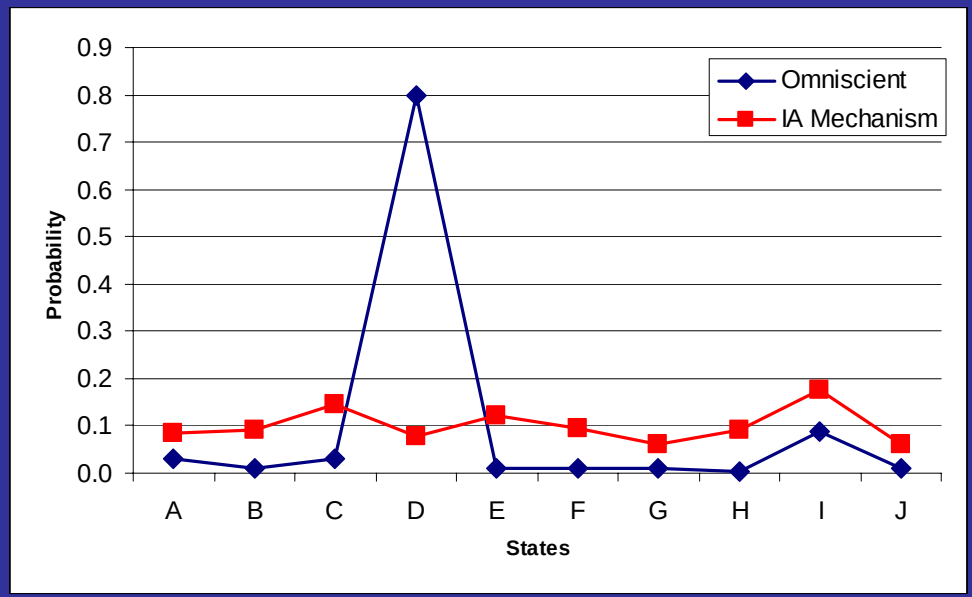


Experiment 4, Period 17
2 Players Aggregated

results

comparison to omniscient probability

Kullback-Leibler = 1.606

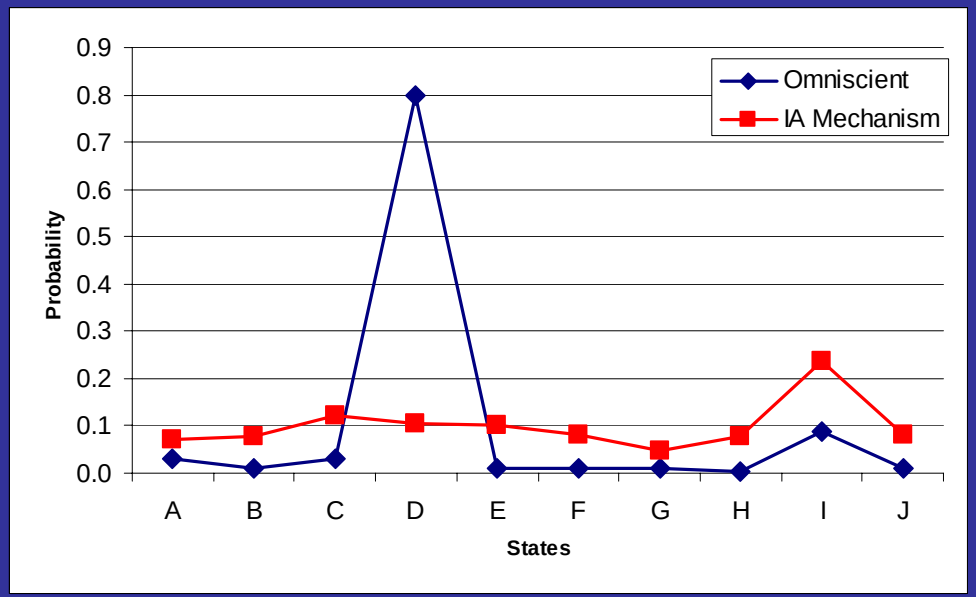


Experiment 4, Period 17
3 Players Aggregated

results

comparison to omniscient probability

Kullback-Leibler = 1.362

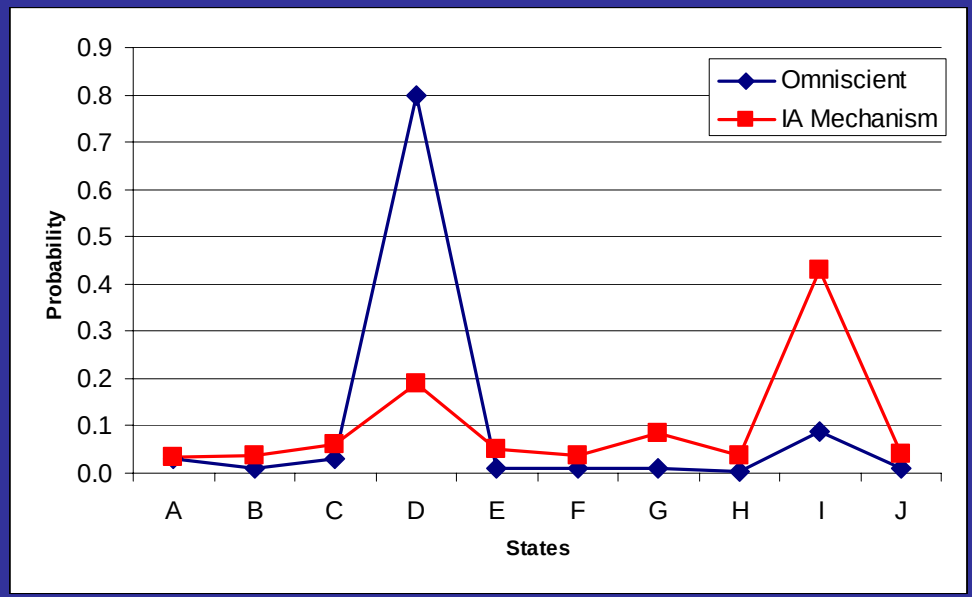


Experiment 4, Period 17
4 Players Aggregated

results

comparison to omniscient probability

Kullback-Leibler = 0.905

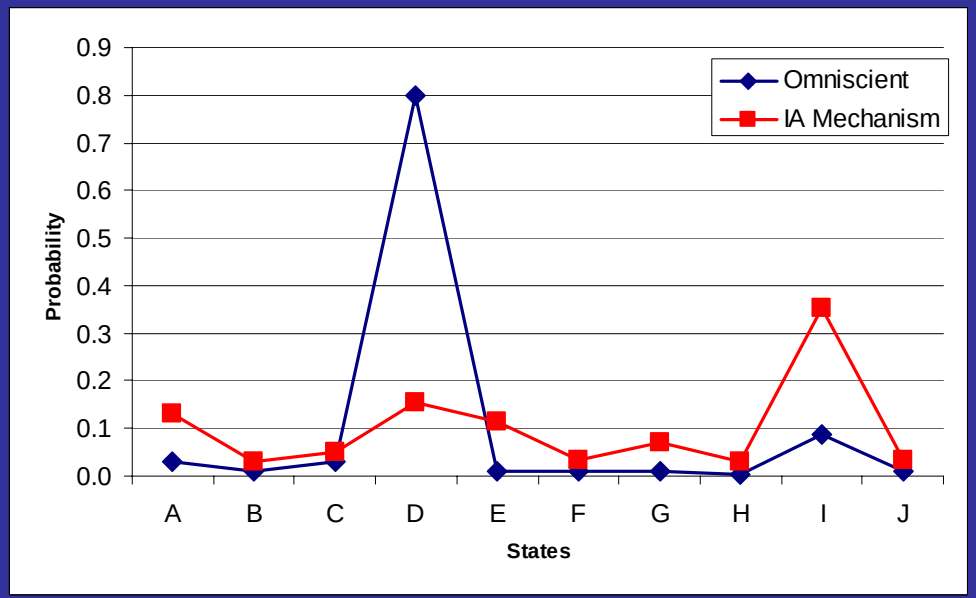


Experiment 4, Period 17
5 Players Aggregated

results

comparison to omniscient probability

Kullback-Leibler = 1.042

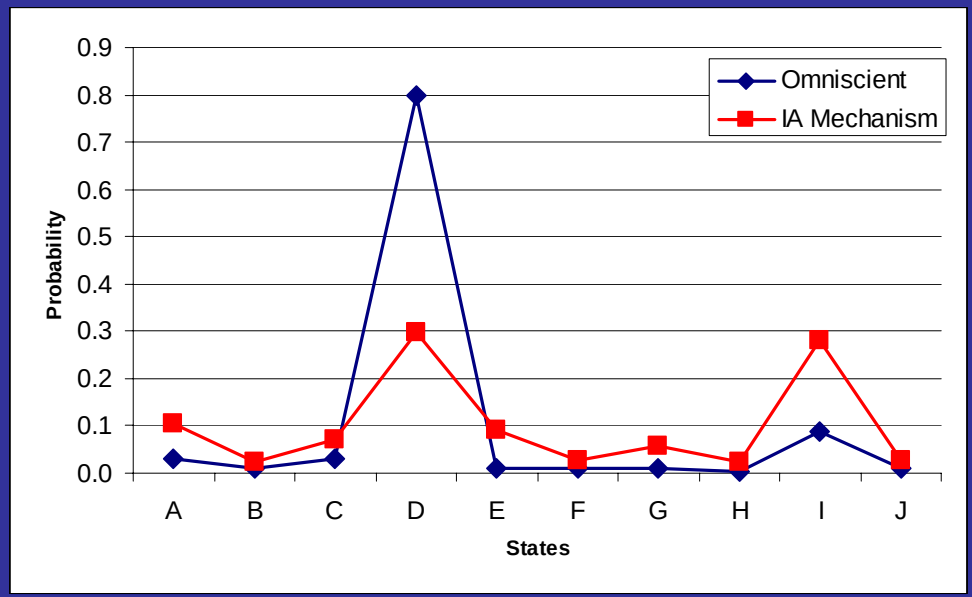


Experiment 4, Period 17
6 Players Aggregated

results

comparison to omniscient probability

Kullback-Leibler = 0.550

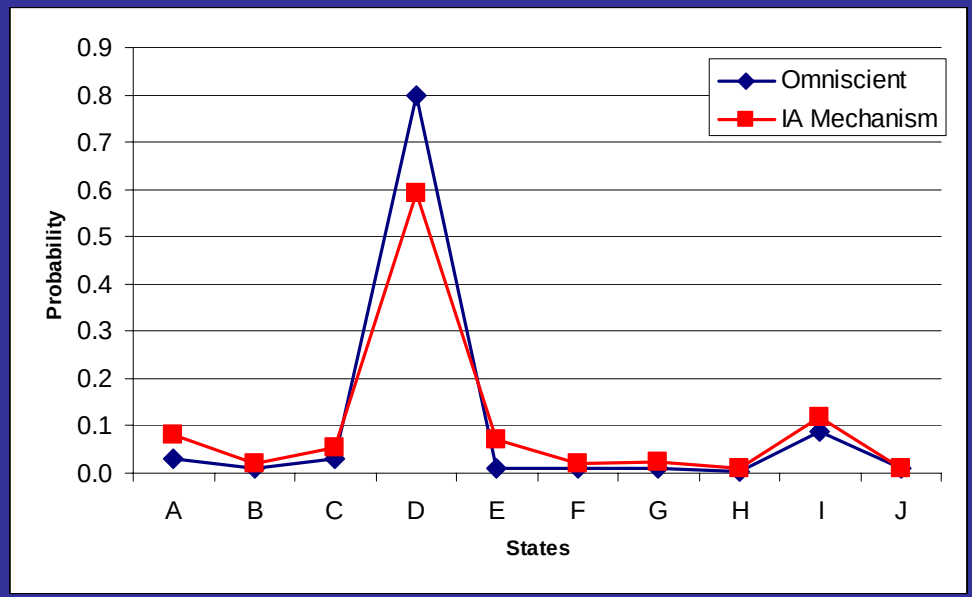


Experiment 4, Period 17
7 Players Aggregated

results

comparison to omniscient probability

Kullback-Leibler = 0.120

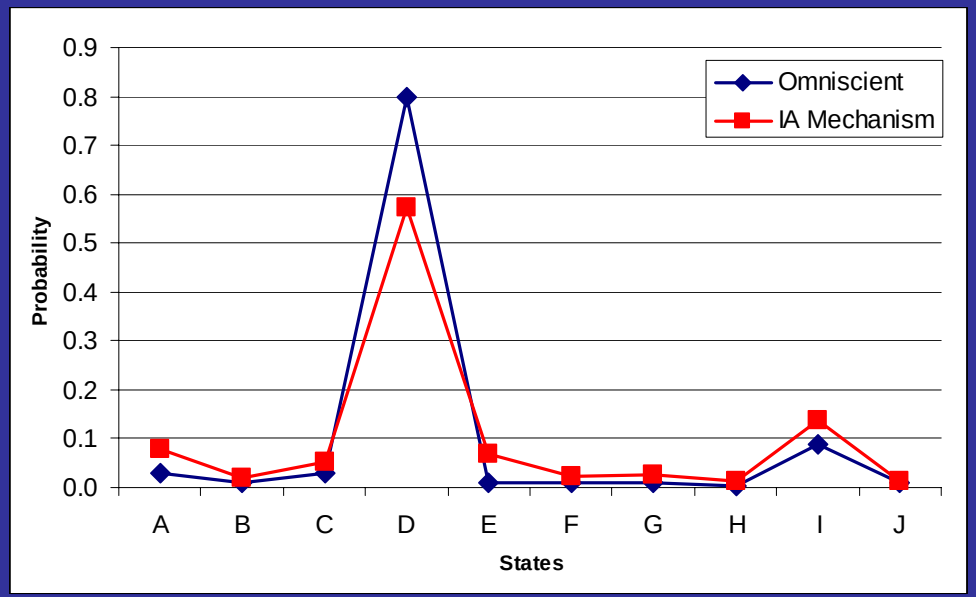


Experiment 4, Period 17
8 Players Aggregated

results

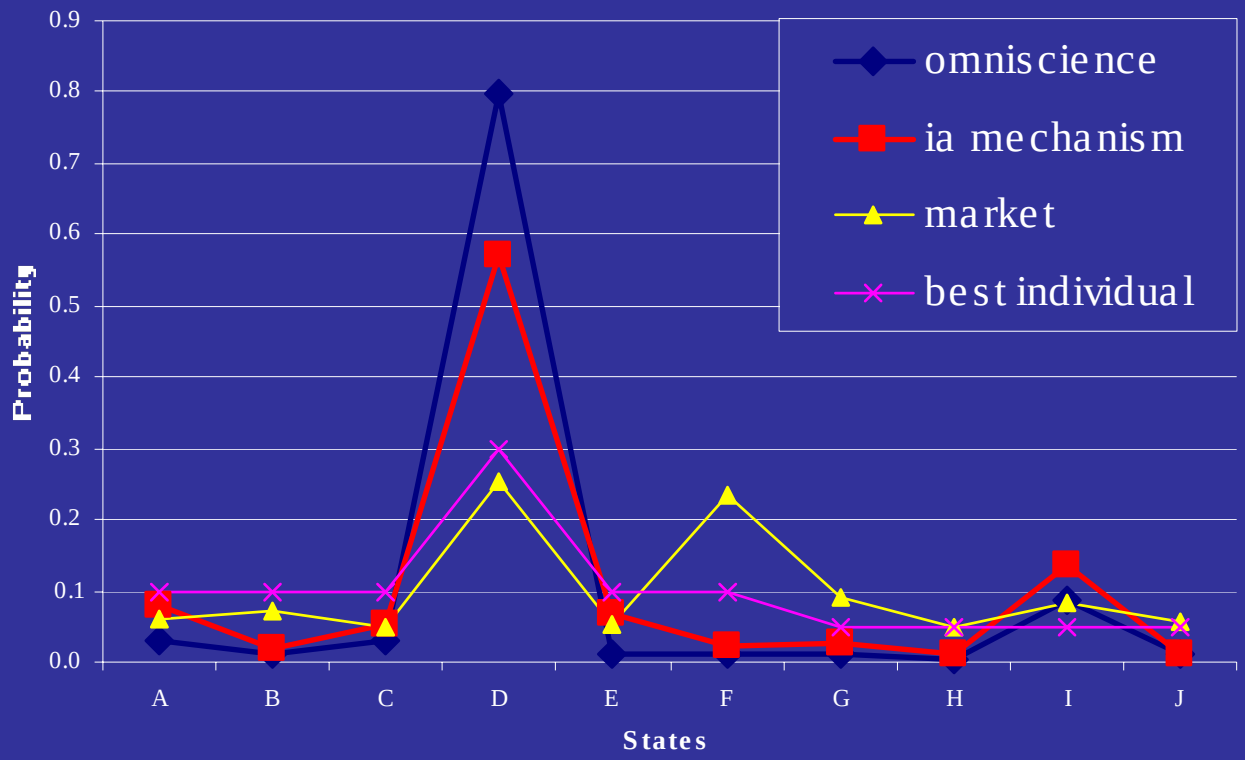
comparison to ominiscient probability

Kullback-Leibler = 0.133



Experiment 4, Period 17
9 Players Aggregated

overall performance



better than the best!

predicting in the real world

(as opposed to the laboratory)

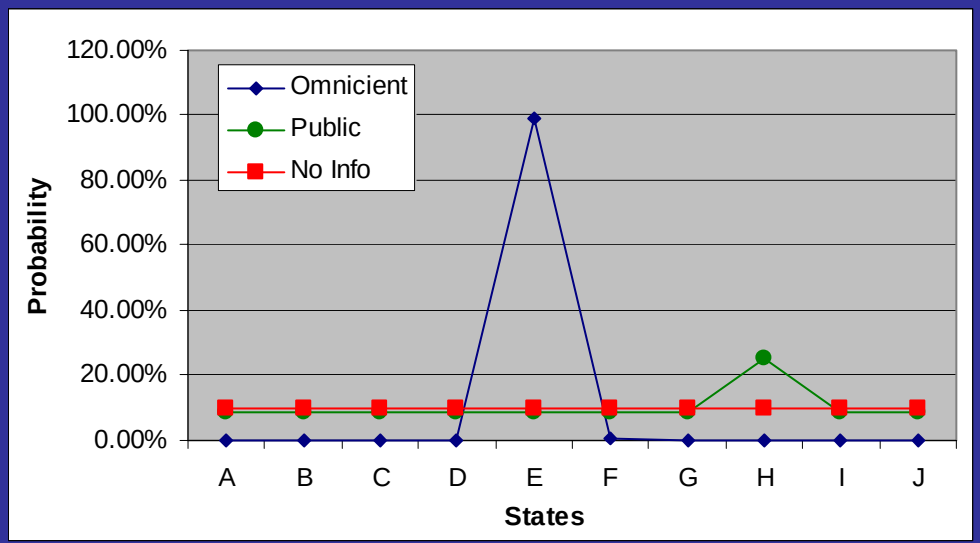
public versus private information

how do we learn which one is which?

and how do eliminate the bias created by

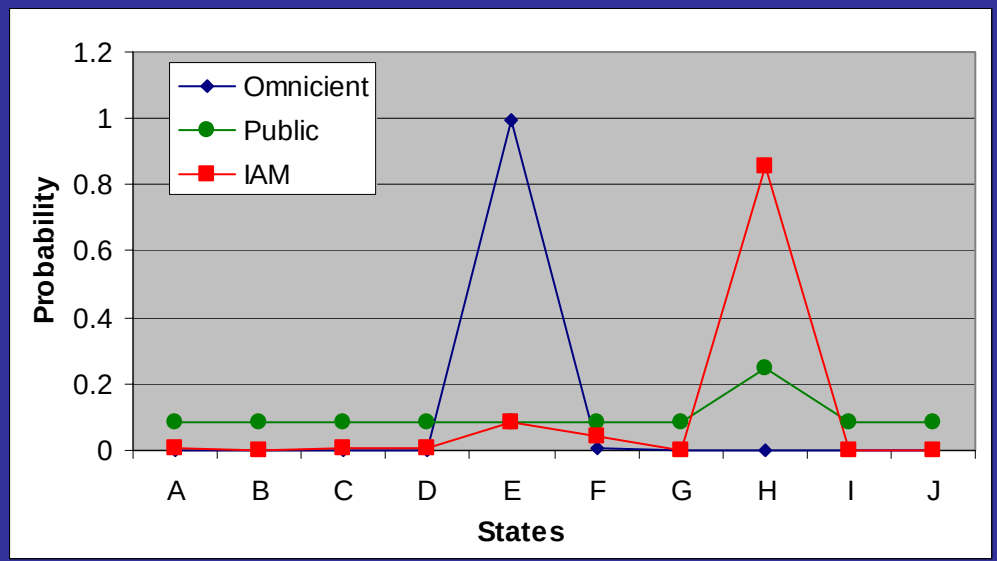
public information?

public information biases



Public Info Experiment 3, Period 9
No Information

public information biases



Public Info Experiment 3, Period 9
9 Players Aggregated

a matching game

payoffs structured such that they elicit:

a) truth revelation – as before

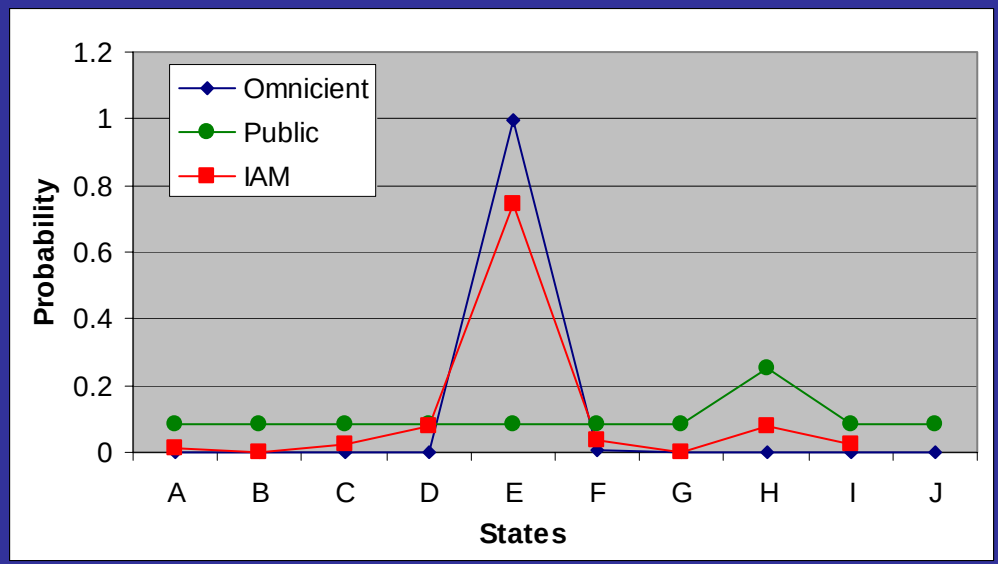
b) good guesses of what others know

matching game

HE \ SHE	B	O
B	3,1	0,0
O	0,0	1,3

correcting for public information

Kullback-Leibler = 0.268



Public Info Experiment 3, Period 9
10 Players Aggregated

for more details

http://www.hpl.hp.com/shl