Informed Truthfulness for Multi-Task Peer Prediction

Victor Shnayder, Arpit Agarwal, Rafael Frongillo, David C. Parkes

Nov 3, 2016; HCOMP'16 Foundations Workshop

Let’s talk about crowdsourcing

Task: How many cows in this image?

Count: 23

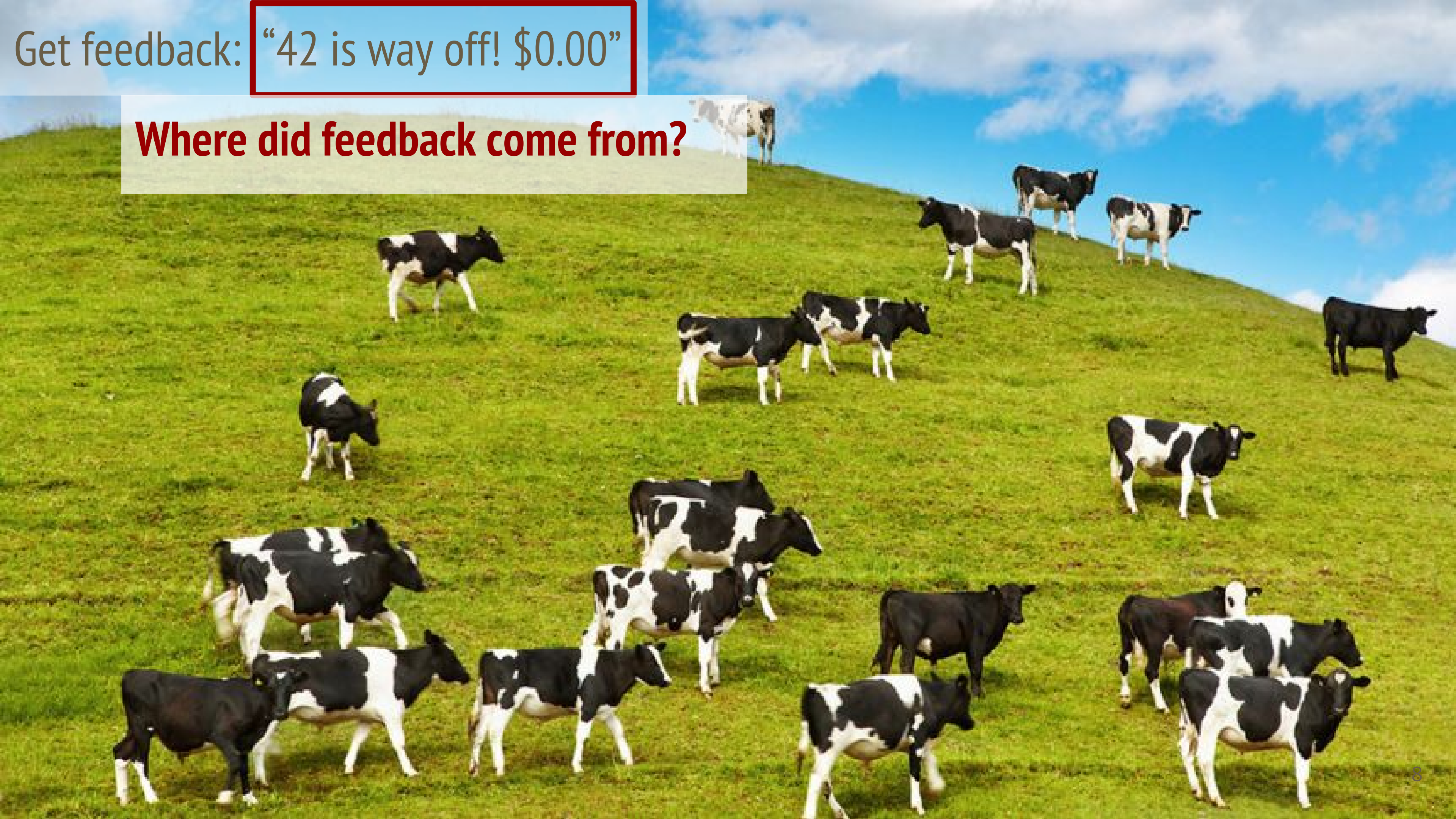
… or just guess: 42

Get feedback: “23 is good. $0.02”

Get feedback: "42 is way off! $0.00"

Get feedback: “42 is way off! $0.00”

Where did feedback come from?




Get feedback: “42 is way off! $0.00”
Where did feedback come from?
Gold standard: professional cow counter said 23

Get feedback: “42 is way off! $0.00”

Where did feedback come from?

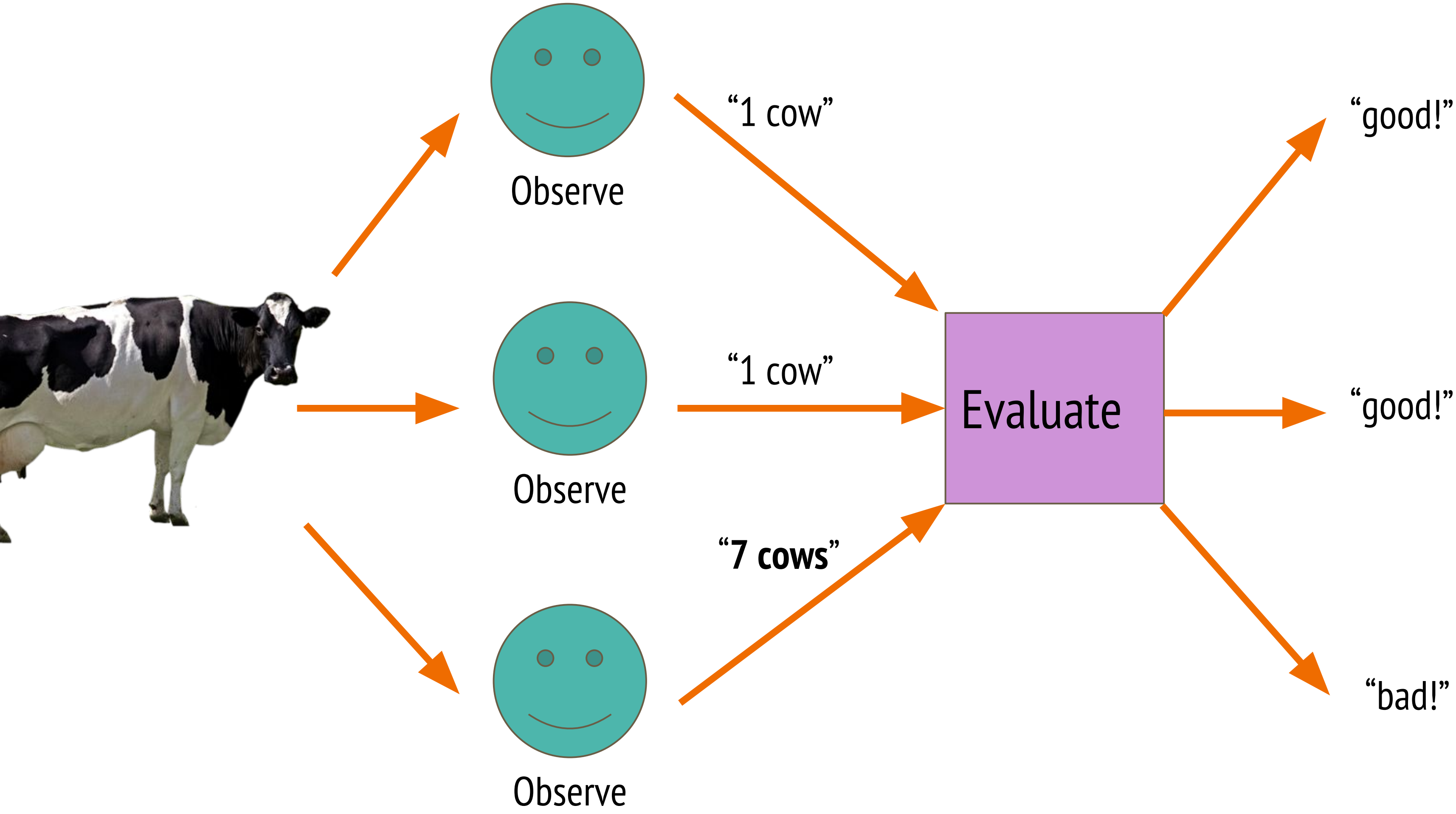
Machine vision: automated cow counter said 20-25

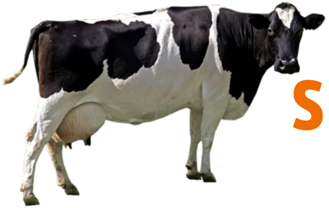
Get feedback: “42 is way off! $0.00”

Where did feedback come from?

Peers: 3 others said
20, 22, 24

Peer prediction: giving feedback based on peer reports


Observe
Observe
Observe
“1 cow”
“1 cow”
“7 cows”
Evaluate
“good!”
“good!”
“bad!”

Applications beyond  s

Gather location-specific info

Image and video labeling

Search result evaluation

Academic peer review

Participatory sensing

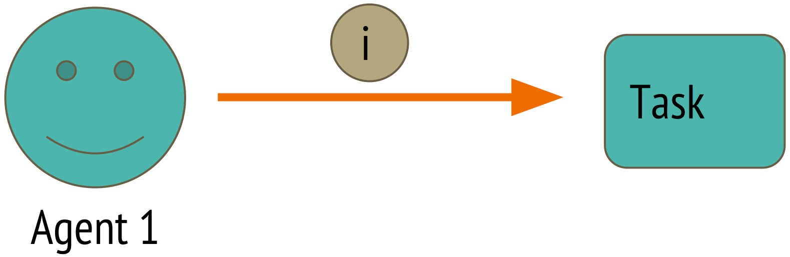
Evaluations for peer assessment in massive courses

Goals:

Ensure truthful equilibrium exists and is attractive

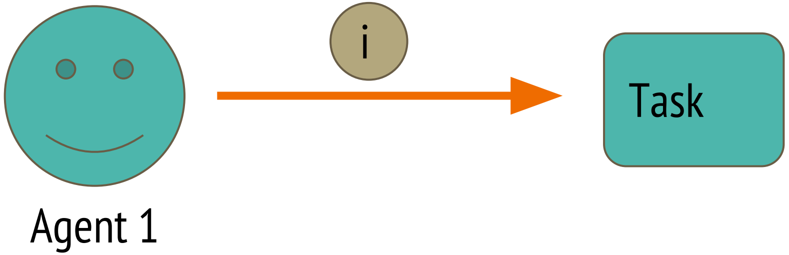
Impossibility results

Task model



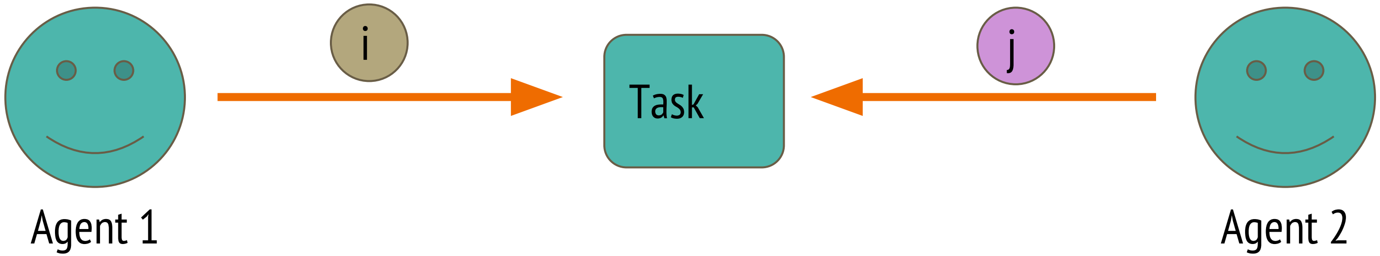
Signal: 1 … n

Task model



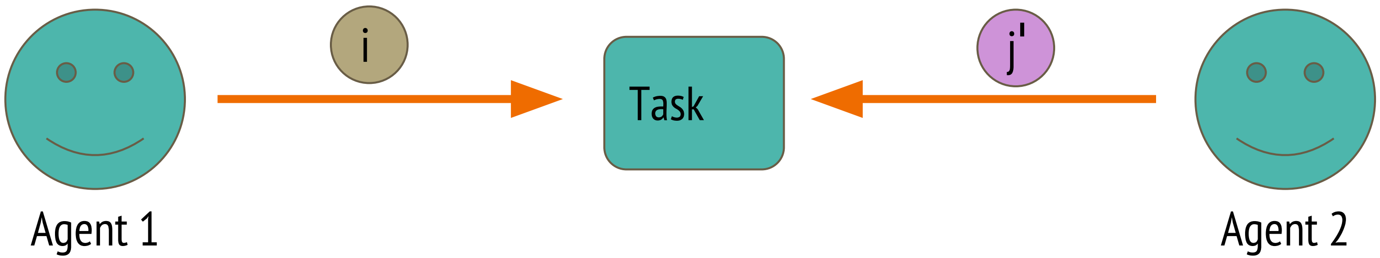
Prior probability P(i)

Task model



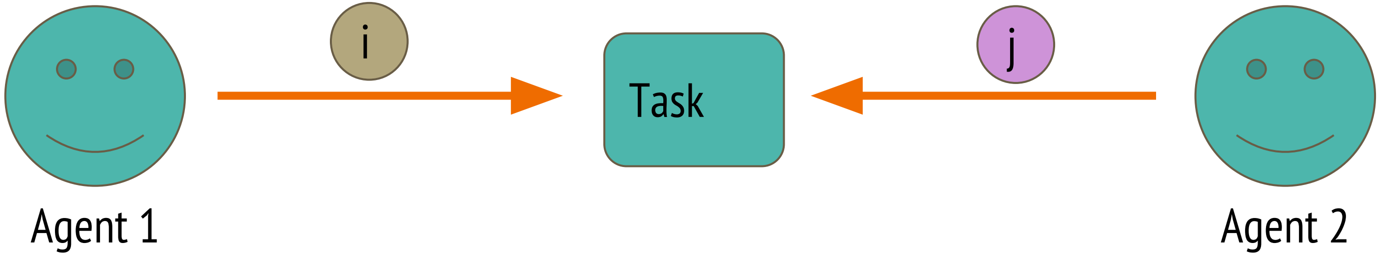
Joint probability: P(i,j)

Task model



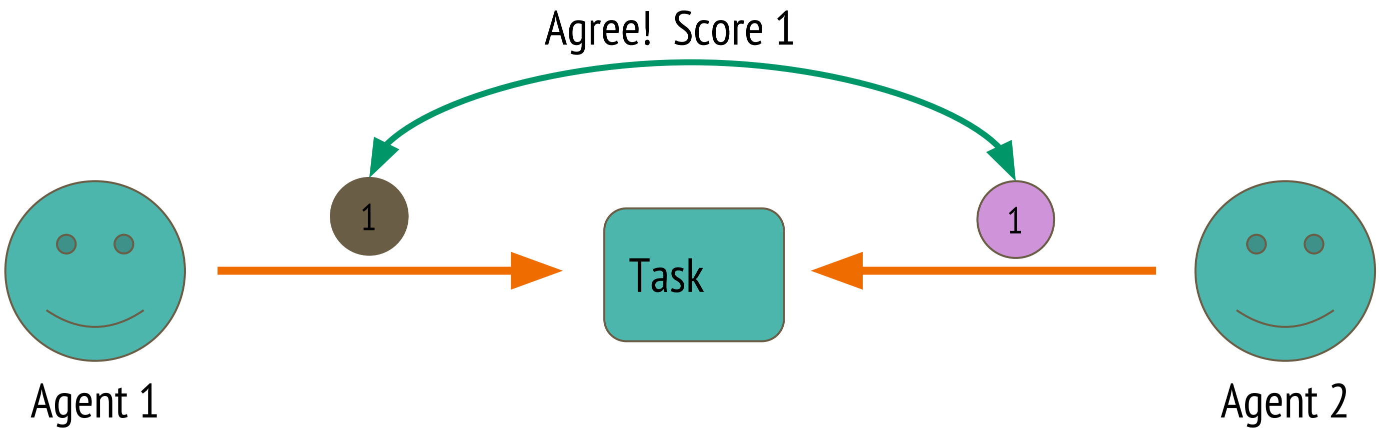
People can misreport!

Task model

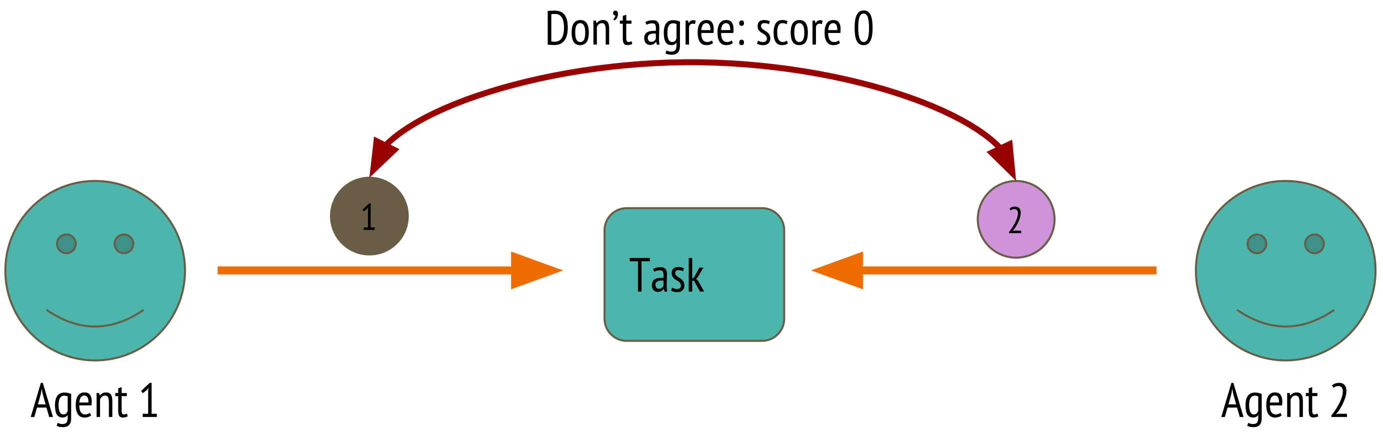


Goal: design scores to encourage effort, truthful reports

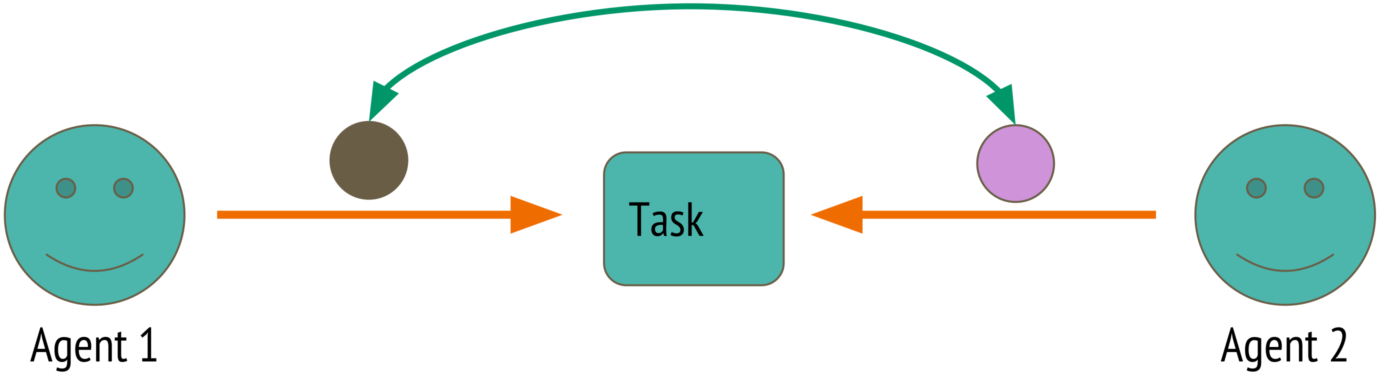
Output agreement (von Ahn, Dabbish '04)



Output agreement

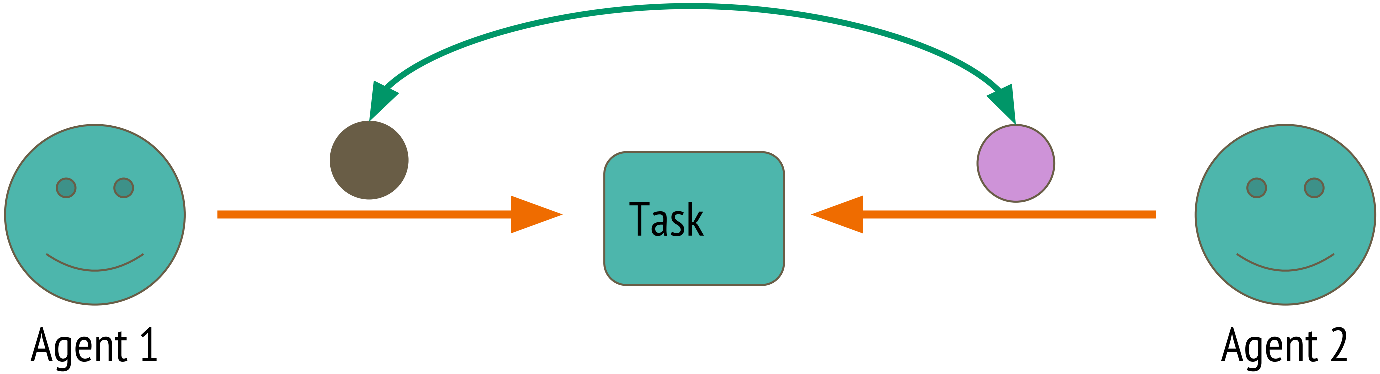


Output agreement



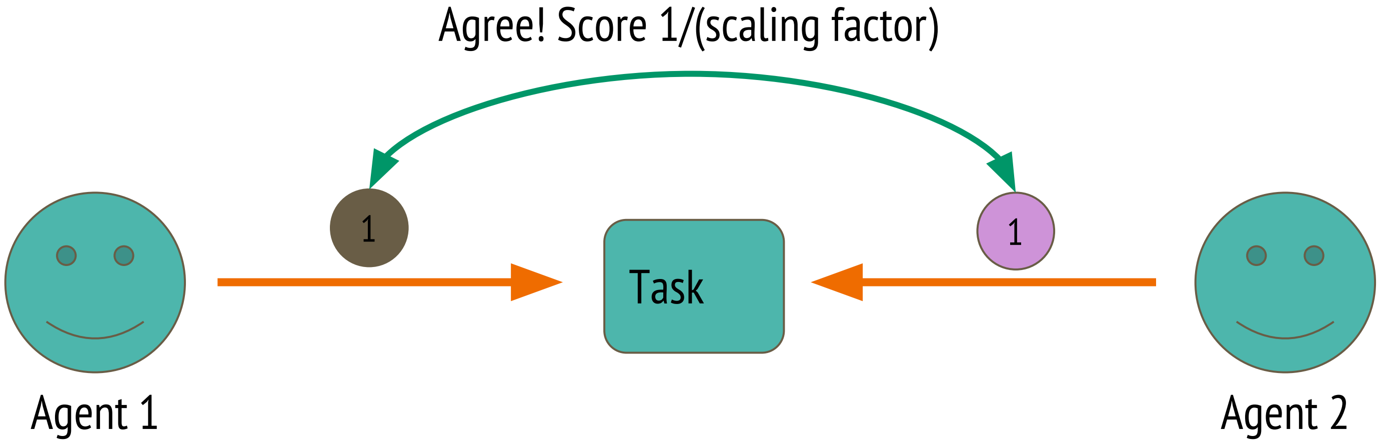
Honest reporting is a correlated equil if my signal predicts yours

Output agreement



To manipulate: all agents always report same thing

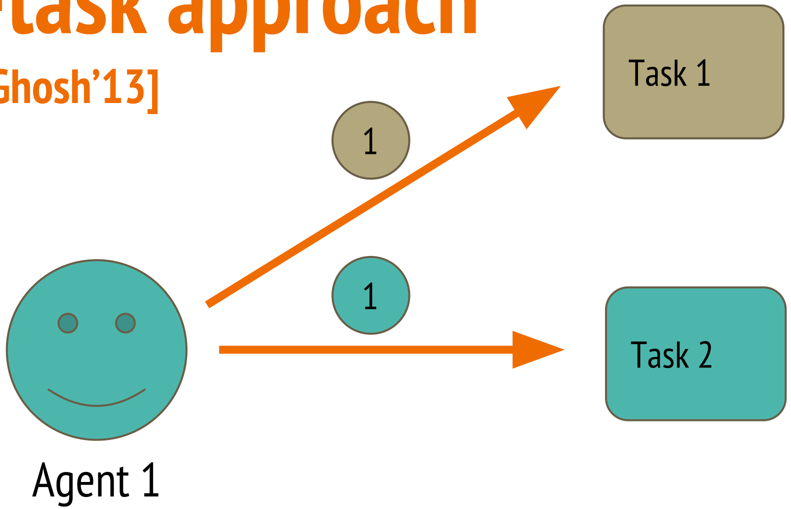
Ensuring truthful reporting is best
[Kamble et. al. ’15, Radanovic et. al. ’16]



Scaling factor learned from reports on many similar tasks.
Truthfulness is an equilibrium, guarantees highest payoff.

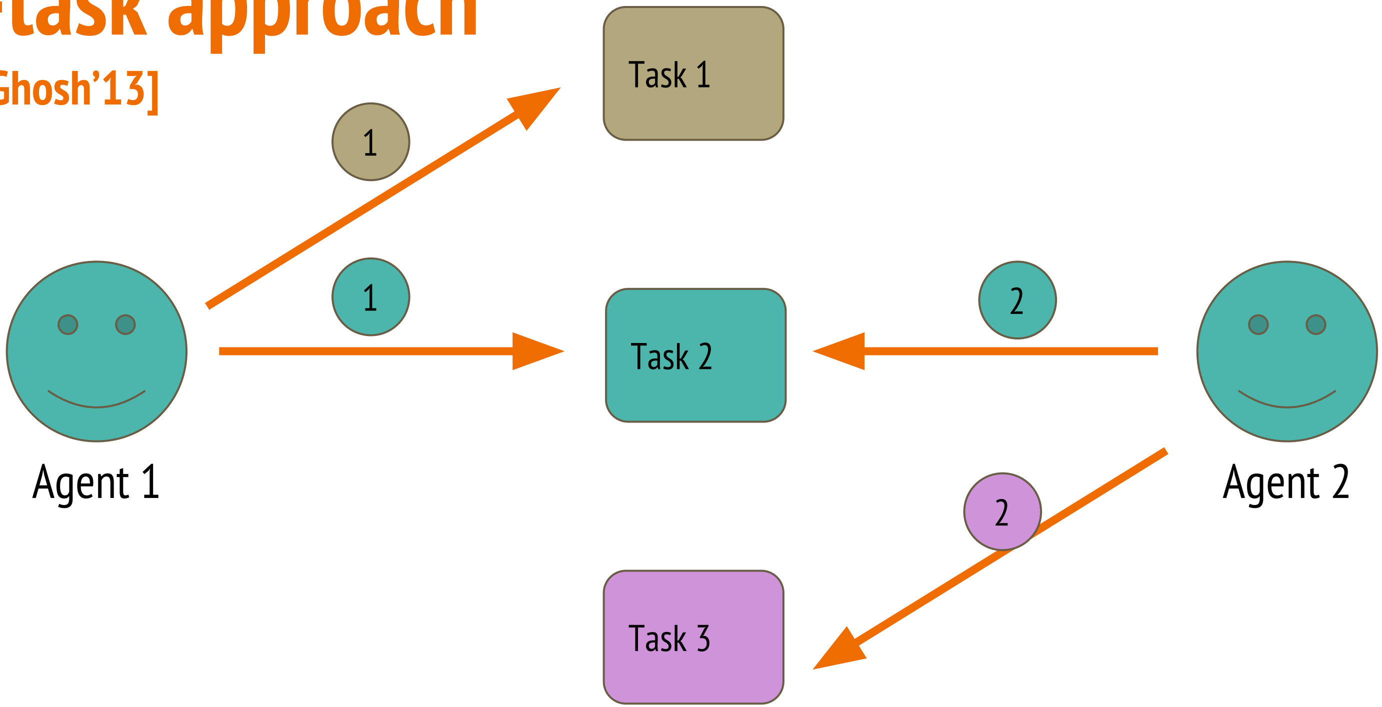
Multi-task approach

[Dasgupta-Ghosh’13]

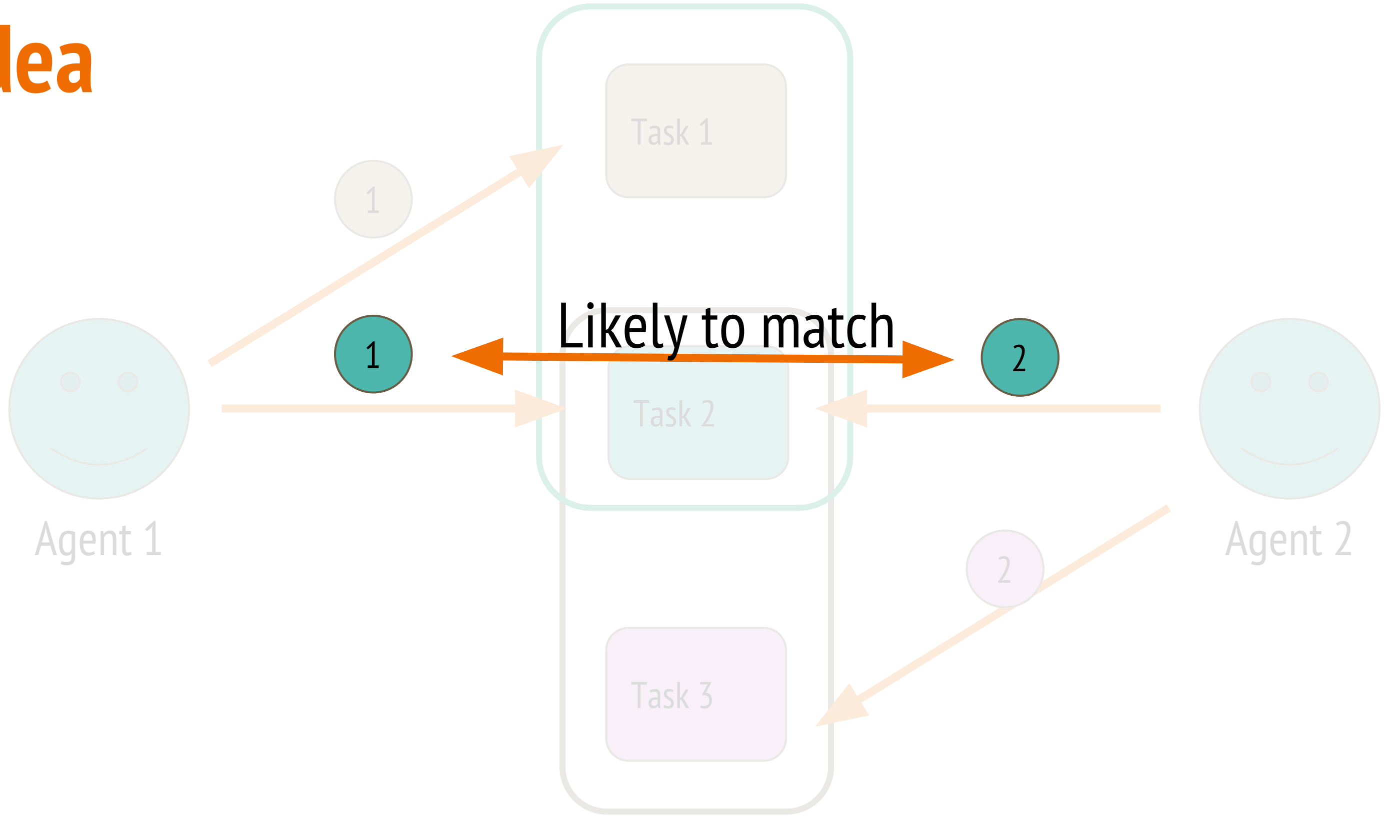


Multi-task approach

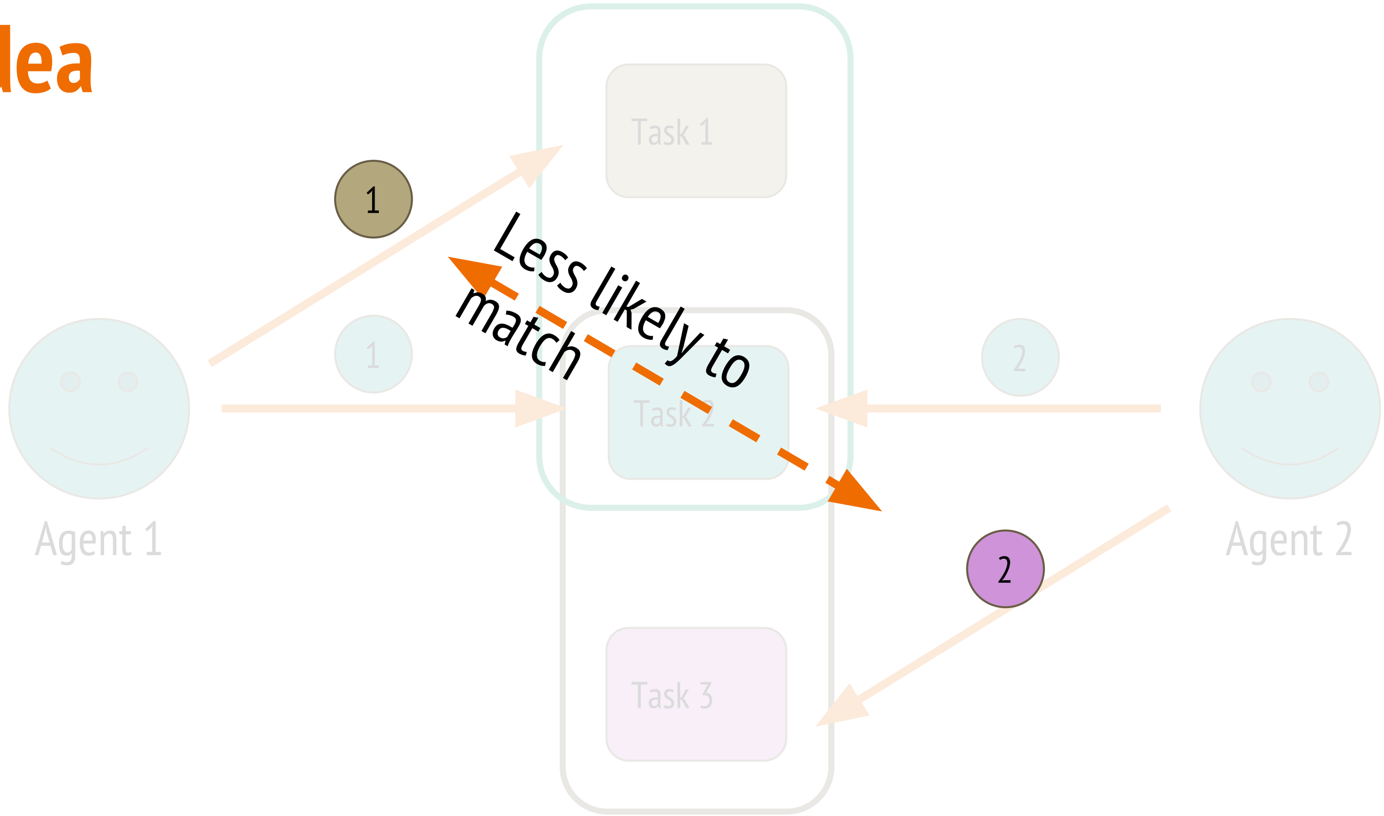
[Dasgupta-Ghosh'13]



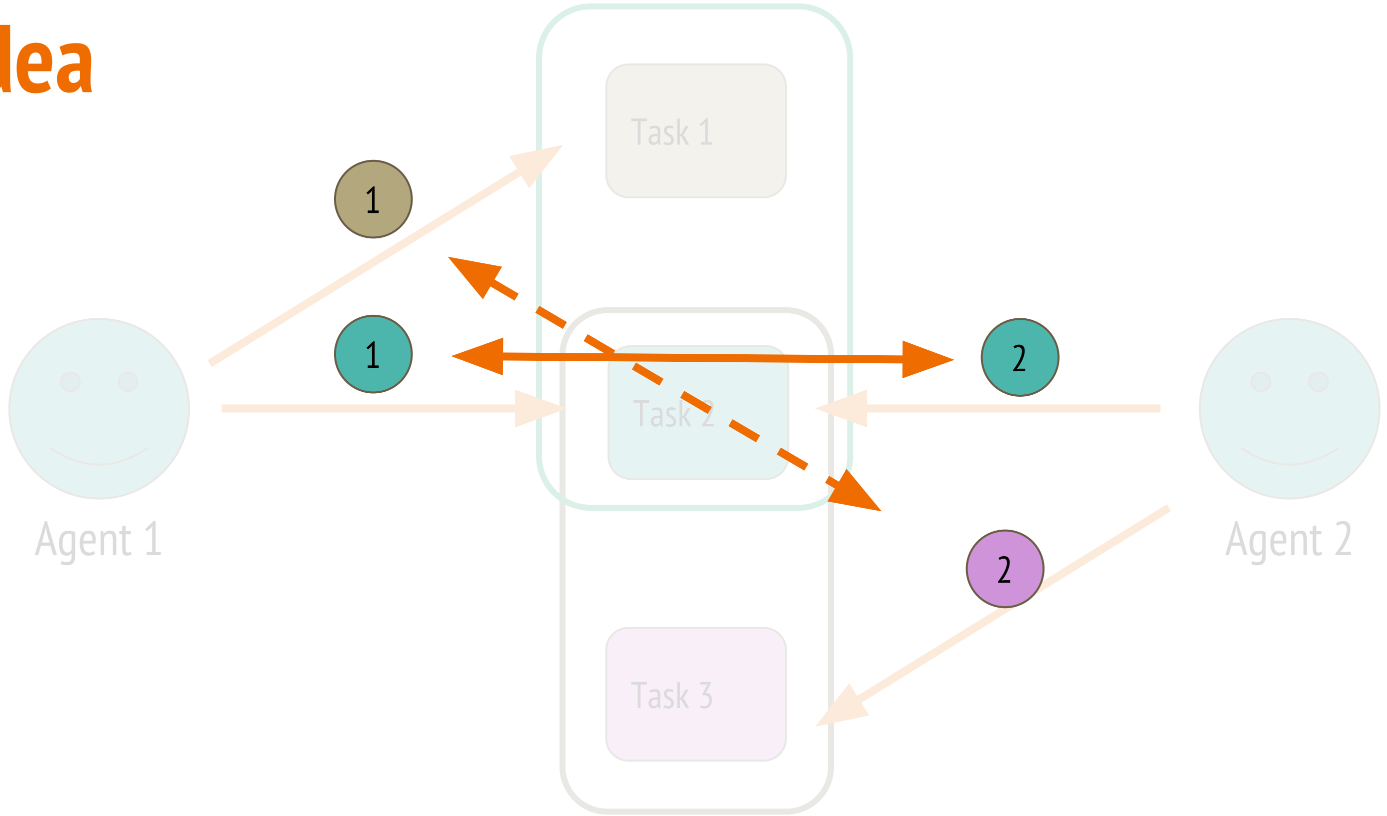
Key idea



Key idea

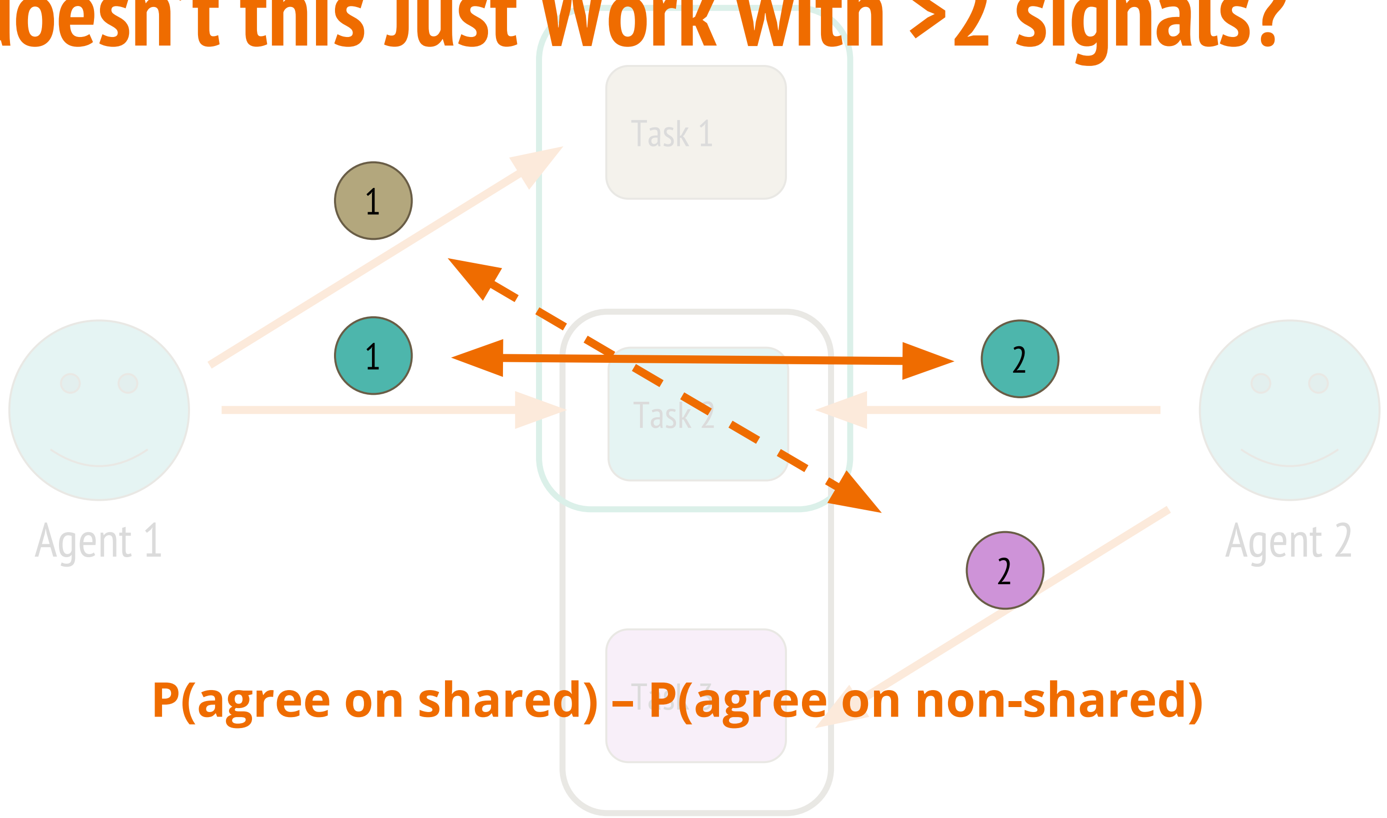


Key idea



Reward matching on shared tasks.
Punish matching on non-shared tasks.

Why doesn't this Just Work with >2 signals?



P(agree on shared) – P(agree on non-shared)

Our multi-signal mechanism: Correlated Agreement

1. Split tasks into shared and non-shared.
2. Score = (agree on shared) – (agree on non-shared)

“Agree” when reports aren’t equal, but positively correlated.

Constant or random reporting still has expected score 0

The *Correlated Agreement* mechanism

1. **Informed truthful**—being truthful is optimal, better than constant or random reports
2. Works with minimal information with few tasks
3. Works with no information with many tasks

Connection to information theory

Truthful score ~ $D_{TV}(P(\cdot,\cdot) - P(\cdot)P(\cdot))$

More agreement ⇒ higher scores.

Other rules correspond to different distance functions.
[Kong-Schoenebeck ‘16]

Open questions

Is peer prediction practical as primary incentive? When?

Combine peer prediction with other incentive models in a single system?

Heterogeneous agents

Non-binary effort models

Non-random task assignment (e.g. maps)

Unintended correlated “signals”

Thank you!

shnayder@post.harvard.edu

Extra slides

Setup

Agents 1, 2, tasks k

Signals i,j (require effort)

Shared tasks, agent 1 tasks, agent 2 tasks

Signal prior P(i), joint P(i,j)

Strategies: F, G probability of reporting r given signal i.

- Informed strategy: depend on the signal somehow
- Truthful strategy: F^*

Solution concepts

E(F, G): expected payment for a shared task

(Strict) Proper: $E(F^*, G^*) \geq E(F, G^*)$, for all $F \neq F^*$

Strong-truthful: $E(F^*, G^*) \geq E(F, G)$, for all F, G (if expected payment tied, F and G must be permutations)

Informed-truthful: $E(F^*, G^*) \geq E(F, G)$, for all F, G (if expected payment tied, F and G must be informed)

A useful matrix

$$\Delta_{ij} = P(i, j) - P(i)P(j)$$

$$\text{If } \Delta_{ij} > 0 \text{ then } P(j|i) > P(j)$$

Example Delta

Prior: $P(i) : [0.55; 0.45]$

Joint: $P(i,j) : \begin{bmatrix} 0.4 & 0.15 \\ 0.15 & .3 \end{bmatrix}$

Delta: $\Delta \approx \begin{pmatrix} 0.1 & -0.1 \\ -0.1 & 0.1 \end{pmatrix}$

Sign(Delta): $\begin{pmatrix} + & - \\ - & + \end{pmatrix}$

Scoring matrix

$$\Delta = \begin{pmatrix} + & + & - \\ + & + & - \\ - & - & + \end{pmatrix} \qquad S = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Expected payment

$$\begin{aligned} E(F,G) &= \sum_{i,j} P(i,j)S(F_i,G_j) - P(i)P(j)S(F_i,G_j) \\ &= \sum_{i,j} (P(i,j) - P(i)P(j))S(F_i,G_j) \\ &= \sum_{i,j} \Delta_{ij} S(F_i,G_j). \end{aligned}$$

(Lemma: can restrict to deterministic strategies)

Key theorem

The Correlated Agreement mechanism is informed truthful for all* models.

Proof sketch

$$E(F,G) = \sum_{i,j} \Delta_{ij} S(F_i, G_j)$$

$$\Delta = \begin{pmatrix} + & + & - \\ + & + & - \\ - & - & + \end{pmatrix} \qquad S = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Other results

Detail-free version of mechanism: learn scoring matrix from reports

Correlated Agreement is maximal among large class of mechanisms that use this scoring structure

Much simpler analysis of Dasgupta-Ghosh'13 mechanism

Binary mechanism (Dasgupta-Ghosh '13):

1. Split tasks into shared and non-shared.
2. Score: (agree on shared) – (agree on non-shared)

Expected score: P(agree on shared) – P(agree on non-shared)

Have to agree based on properties of shared task.
Constant or random reporting has expected score = 0