Information Elicitation Sans Verification

Bo Waggoner and Yiling Chen

2013-06-16

Motivation: human computation

Motivation: human computation

Motivation: human computation

amazonmechanical turk
beta
Artificial Artificial Intelligence

CrowdFlower

Goal: design systems for eliciting info

Goal: design systems for eliciting info

Question: How to construct human computation systems?

Goal: design systems for eliciting info

Question: How to construct human computation systems?
Approach: Use *mechanism design*

Mechanism design

Mechanism design:

Construct a game to optimize an objective

Mechanism design

Mechanism design:

Construct a game to optimize an objective

Game: different actions available; set of actions maps to an outcome and payoffs.

Mechanism design

Mechanism design:

Construct a game to optimize an objective

Mechanism design

Mechanism design:

Construct a game to optimize an objective

Our objective: elicit "useful" information

Mechanism design

Mechanism design:

Construct a game to optimize an objective

Our objective: elicit "useful" information

Our constraints:

Mechanism design

Mechanism design:

Construct a game to optimize an objective

Our objective: elicit "useful" information

Our constraints:

1. players may not prefer "useful" responses

Mechanism design

Mechanism design:

Construct a game to optimize an objective

Our objective: elicit "useful" information

Our constraints:

1. players may not prefer "useful" responses
2. designer cannot always verify responses

Mechanism design

Mechanism design:

Construct a game to optimize an objective

Our objective: elicit "useful" information

Our constraints:

1. players may not prefer "useful" responses
2. designer cannot always verify responses

Our name for this setting:

Information Elicitation Without Verification (IEWV)

Agenda

Plan:

1. Formally define the setting,
   identify limitations of prior work.

Agenda

Plan:

1. Formally define the setting,
   identify limitations of prior work.
2. Prove impossibility results on the setting;
   demonstrate difficulty of overcoming
   limitations.

Agenda

Plan:

1. Formally define the setting,
   identify limitations of prior work.
2. Prove impossibility results on the setting;
   demonstrate difficulty of overcoming
   limitations.
3. Propose new mechanism that overcomes some
   limitations, avoids some impossibilities.

Outline

Information elicitation without verification

Formal setting and prior work

Impossibility results for IEWV

Output agreement mechanisms

Outline

Information elicitation without verification

Formal setting and prior work

Impossibility results for IEWV

Output agreement mechanisms

Setting

Game of information elicitation without verification:

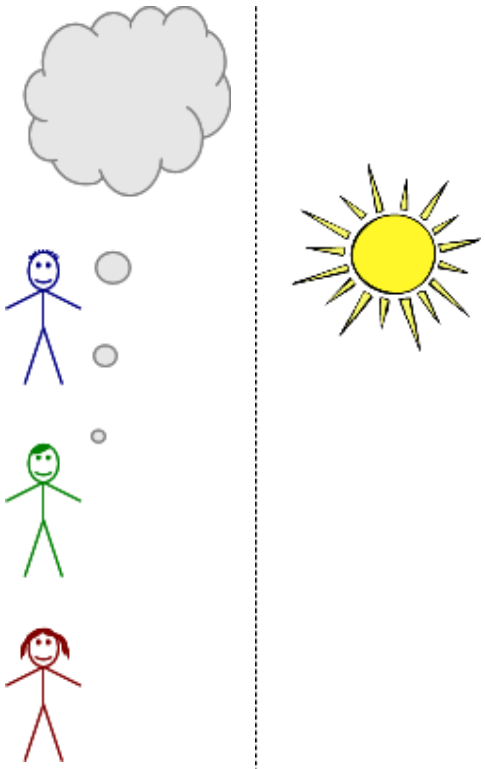
Setting

Game of information elicitation without verification:

prior

Setting

Game of information elicitation without verification:

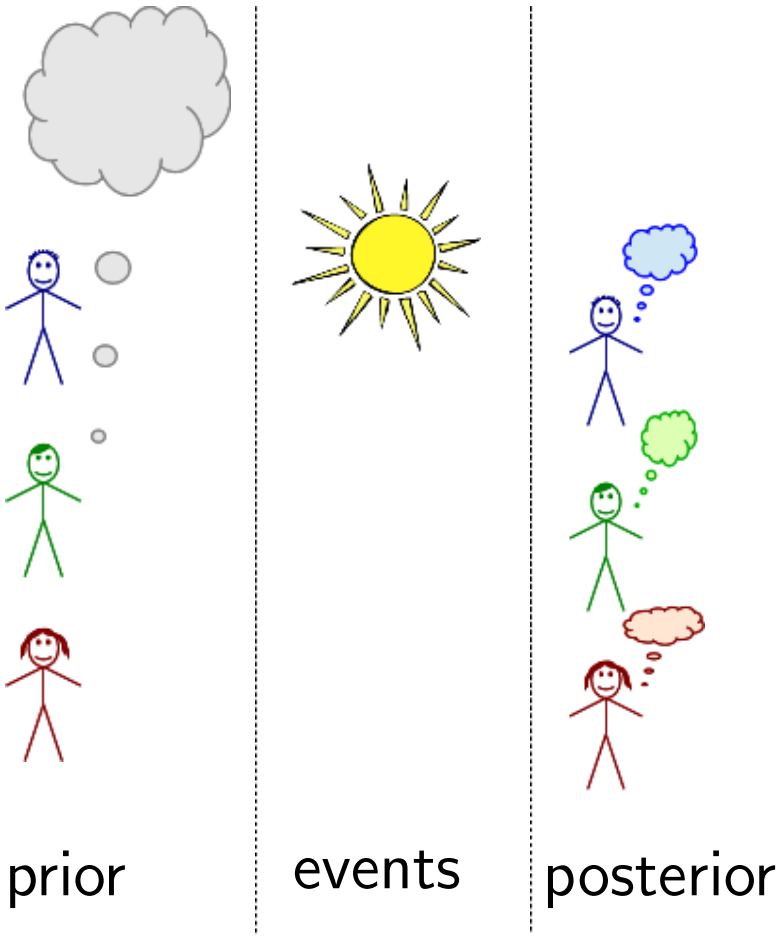


prior

events

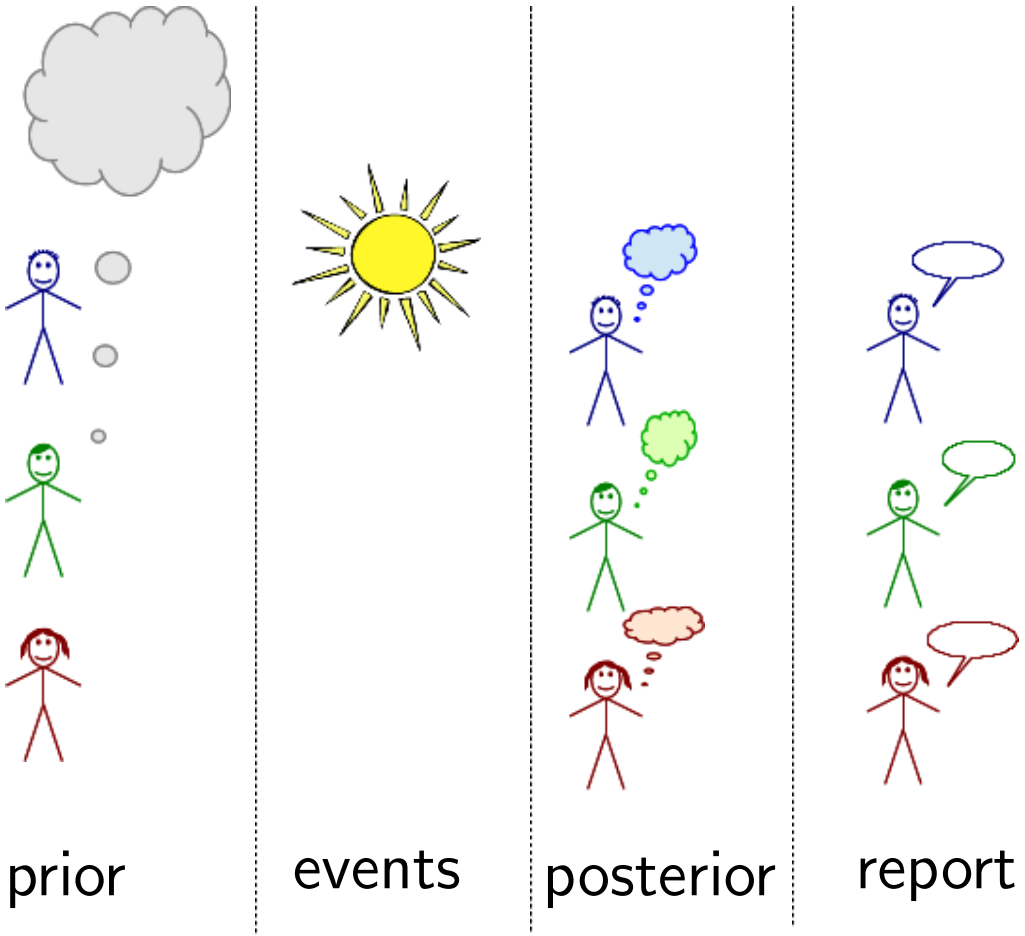
Setting

Game of information elicitation without verification:



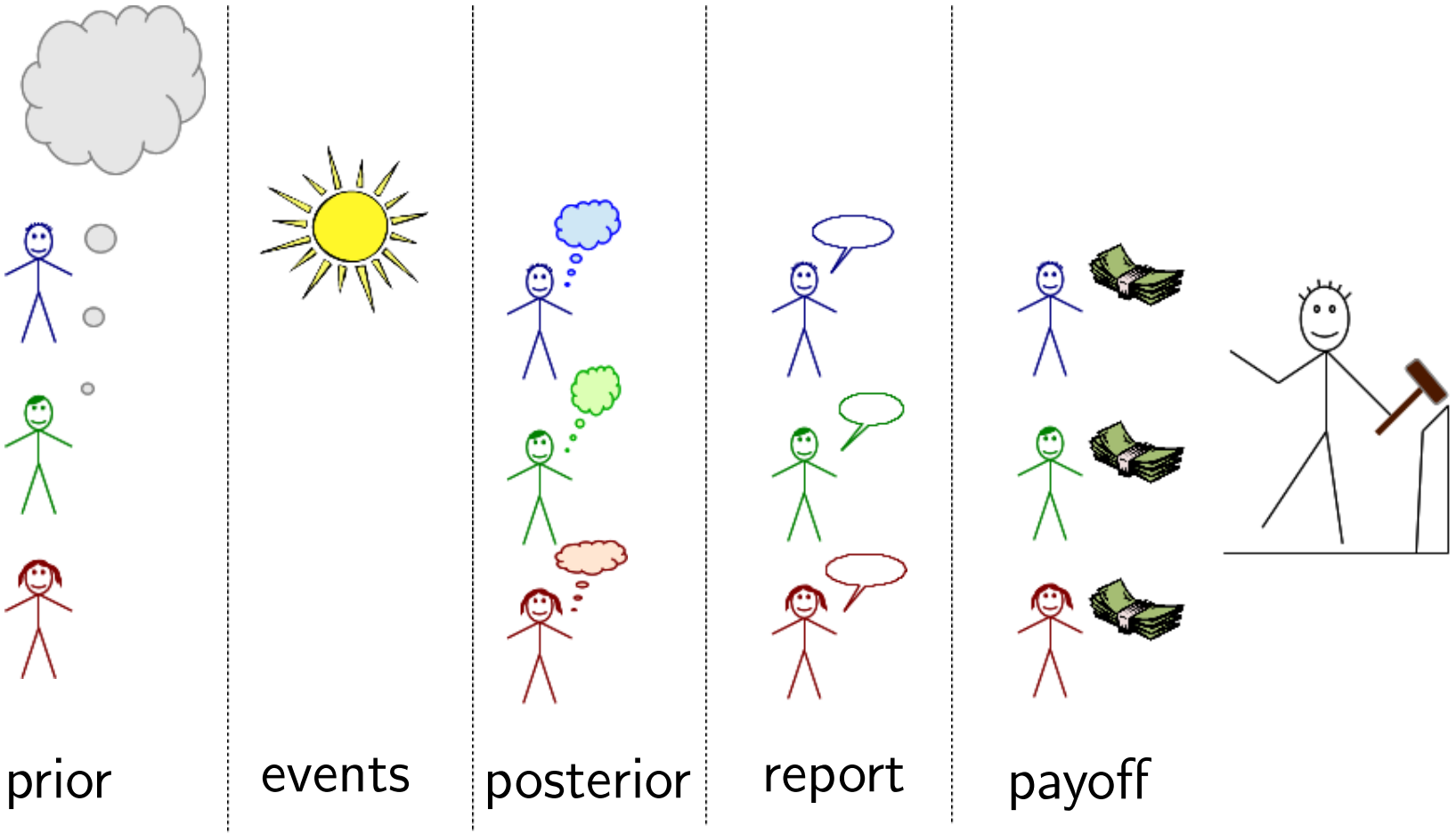
Setting

Game of information elicitation without verification:



Setting

Game of information elicitation without verification:



Prior work: themes

Prior work: various mechanisms for instances of this setting:

Prior work: themes

Prior work: various mechanisms for instances of this setting:

- Peer prediction (Miller, Resnick, Zeckhauser 2005)

Prior work: themes

Prior work: various mechanisms for instances of this setting:

- Peer prediction (Miller, Resnick, Zeckhauser 2005)
- Bayesian truth serum (Prelec 2004)

Prior work: themes

Prior work: various mechanisms for instances of this setting:

- Peer prediction (Miller, Resnick, Zeckhauser 2005)
- Bayesian truth serum (Prelec 2004)
- PP without a common prior, Robust BTS

  (Witkowski, Parkes 2012a,b)

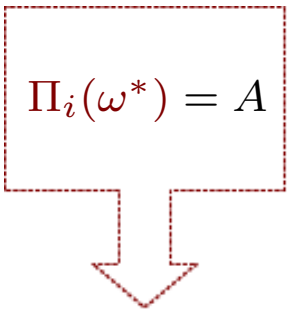
Prior work: themes

Prior work: various mechanisms for instances of this setting:

- Peer prediction (Miller, Resnick, Zeckhauser 2005)
- Bayesian truth serum (Prelec 2004)
- PP without a common prior, Robust BTS (Witkowski, Parkes 2012a,b)
- Collective revelation (Goel, Reeves, Pennock 2009)
  Truthful surveys (Lambert, Shoham 2008)

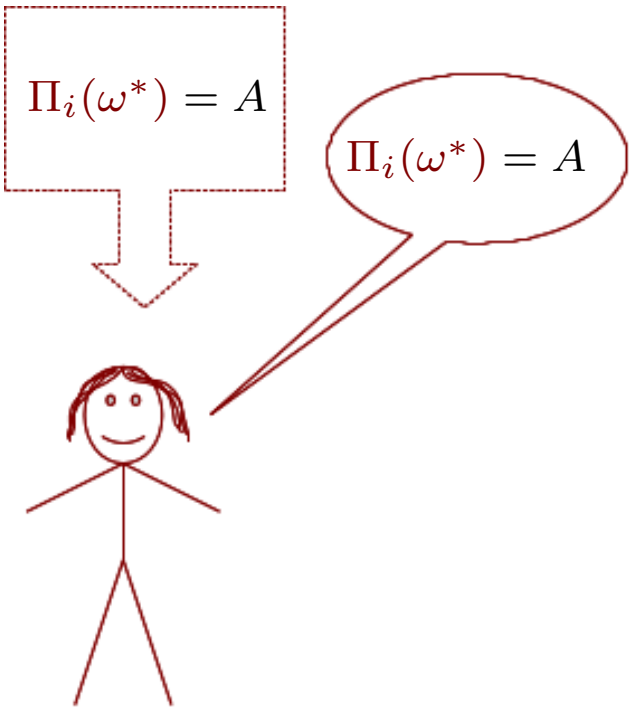
Example: peer prediction

Example: peer prediction



observation

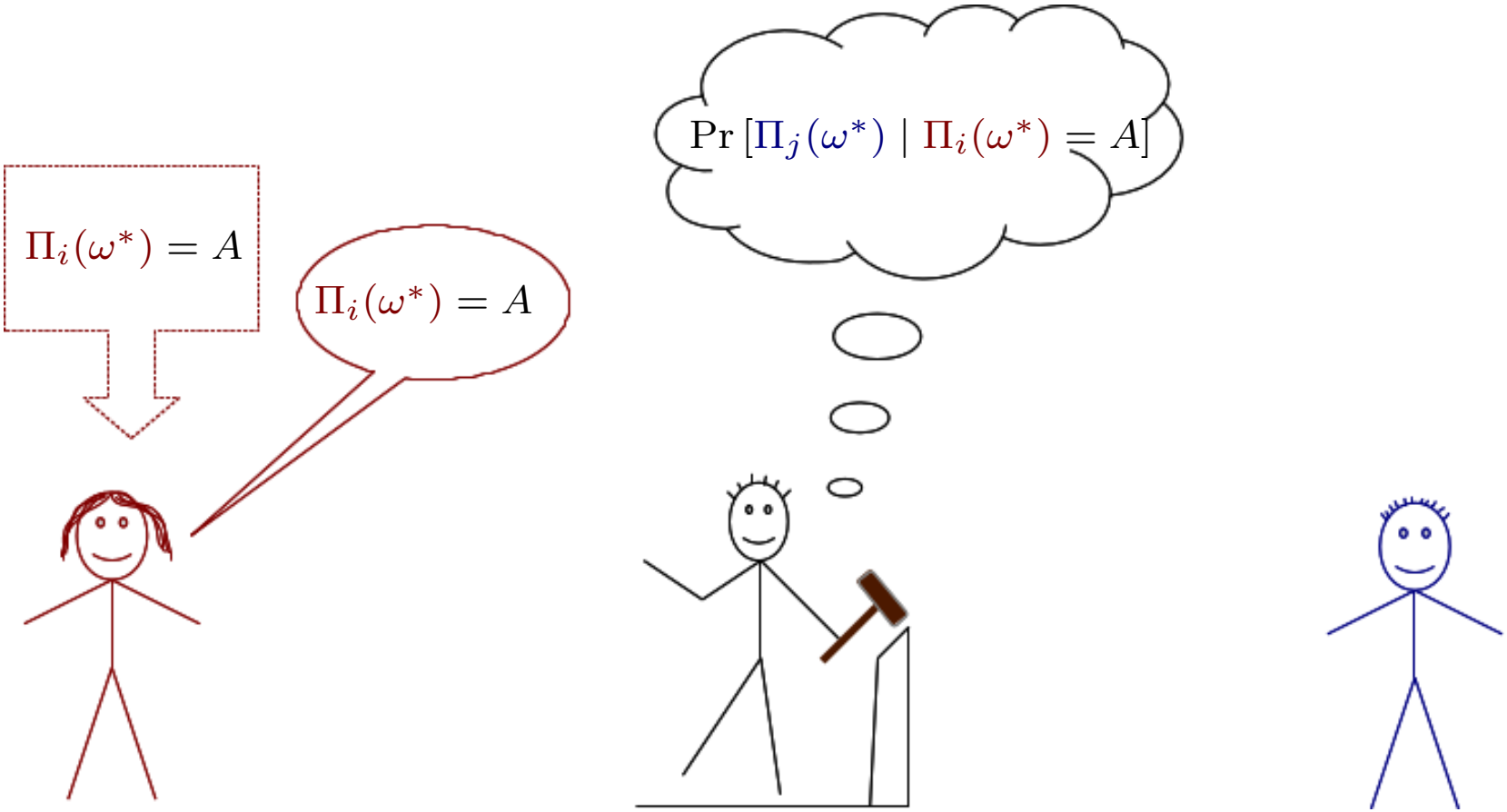
Example: peer prediction



observation

report

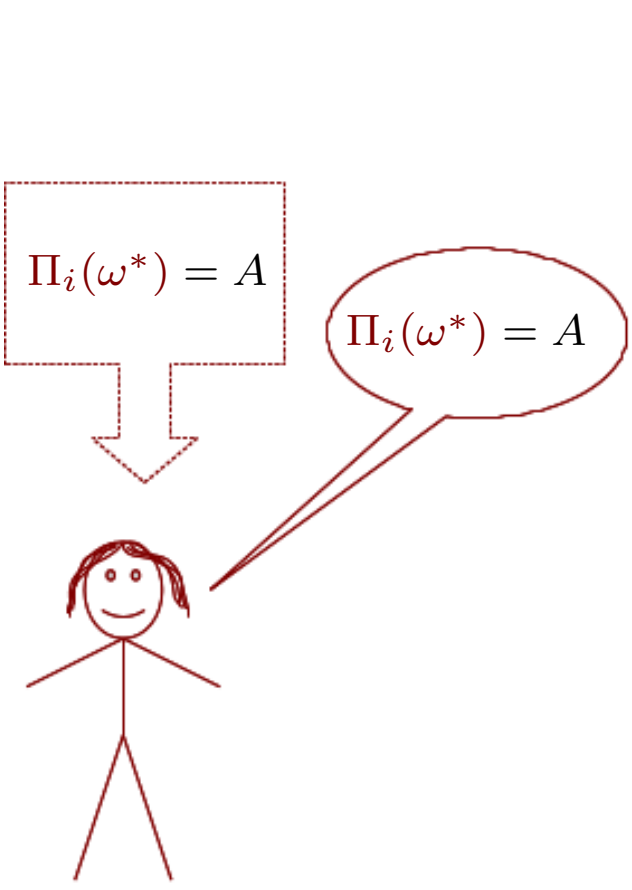
Example: peer prediction



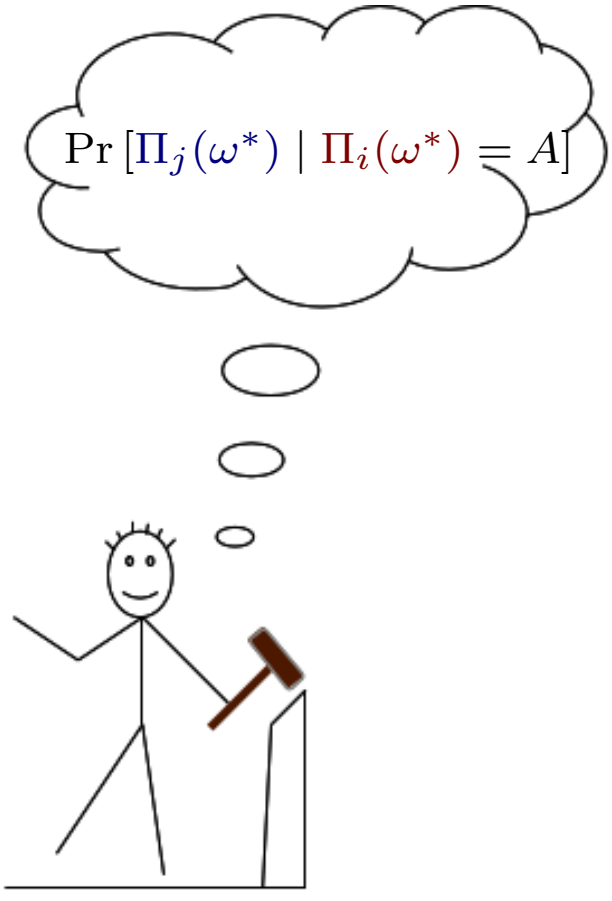
observation report

prediction

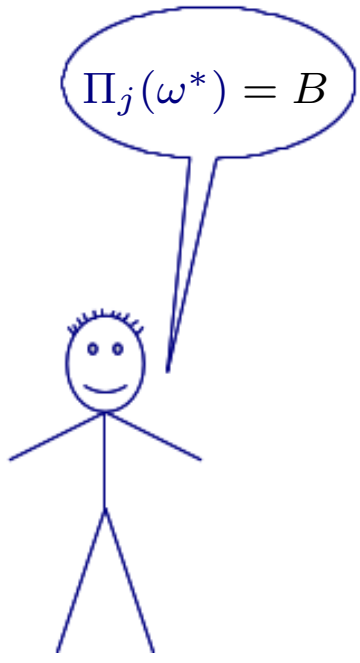
Example: peer prediction



observation



prediction



report

payoff: h a proper scoring rule

$h(\Pr[\Pi_j(\omega^*) \mid \Pi_i(\omega^*) = A], B)$

Prior work: discussion

Prior work: discussion

Limitations of mechanisms in prior work:

- Somewhat complicated to explain

Prior work: discussion

Limitations of mechanisms in prior work:

- Somewhat complicated to explain
- Only applicable in specific settings (*e.g.* elicit signals)

Prior work: discussion

Limitations of mechanisms in prior work:

- Somewhat complicated to explain
- Only applicable in specific settings (*e.g.* elicit signals)
- "Bad" equilibria exist

Prior work: discussion

Limitations of mechanisms in prior work:

- Somewhat complicated to explain
- Only applicable in specific settings (*e.g.* elicit signals)
- "Bad" equilibria exist
- Not detail-free (peer prediction)

Prior work: discussion

Limitations of mechanisms in prior work:

- Somewhat complicated to explain
- Only applicable in specific settings (*e.g.* elicit signals)
- "Bad" equilibria exist
- Not detail-free (peer prediction)
- Restricted domain (all)

Prior work: discussion

Limitations of mechanisms in prior work:

- Somewhat complicated to explain
- Only applicable in specific settings (*e.g.* elicit signals)
- "Bad" equilibria exist
- Not detail-free (peer prediction)
- Restricted domain (all)

Prior work: discussion

Limitations of mechanisms in prior work:

- Somewhat complicated to explain
- Only applicable in specific settings (*e.g.* elicit signals)
- "Bad" equilibria exist
- Not detail-free (peer prediction)
- Restricted domain (all)

Goal: Overcome these limitations.

Prior work: discussion

Limitations of mechanisms in prior work:

- Somewhat complicated to explain
- Only applicable in specific settings (*e.g.* elicit signals)
- "Bad" equilibria exist
- Not detail-free (peer prediction)
- Restricted domain (all)

Goal: Overcome these limitations.
Obstacle: Impossibility results!

Outline

Information elicitation without verification

Formal setting and prior work

Impossibility results for IEWV

Output agreement mechanisms

Existence of uninformative equilibria

Existence of uninformative equilibria

Definition

A strategy is **uninformative** *if it draws a report from the same distribution in every state of the world.*

Existence of uninformative equilibria

Proposition

The following mechanisms for IEWV always have uninformative equilibria:

- *Those with compact action spaces and continuous reward functions;*
- *Those that: (a) are detail-free and (b) always have an equilibrium.*

Existence of uninformative equilibria

Proposition

The following mechanisms for IEWV always have uninformative equilibria:

- *Those with compact action spaces and continuous reward functions;*
- *Those that: (a) are detail-free and (b) always have an equilibrium.*

$\Longrightarrow$ *All mechanisms we know of; all "reasonable" mechanisms.*

Impossibility for truthful equilibria

Impossibility for truthful equilibria

Q: What is "truthful"?

Impossibility for truthful equilibria

Q: What is "truthful"?

A: define a **query** T specifying the truthful response for a given posterior belief.

Impossibility for truthful equilibria

Q: What is "truthful"?

A: define a **query** T specifying the truthful response for a given posterior belief.

truthful strategy: $s_i(\Pi_i(\omega^*)) = T(\Pi_i(\omega^*))$.

Impossibility for truthful equilibria

Q: What is "truthful"?

A: define a **query** T specifying the truthful response for a given posterior belief.

truthful strategy: $s_i(\Pi_i(\omega^*)) = T(\Pi_i(\omega^*))$.
truthful equilibrium: (Given T) one in which each s_i is truthful.

Impossibility for truthful equilibria

Q: What is "truthful"?

A: define a **query** T specifying the truthful response for a given posterior belief.

truthful strategy: $s_i(\Pi_i(\omega^*)) = T(\Pi_i(\omega^*))$.
truthful equilibrium: (Given T) one in which each s_i is truthful.

Theorem

For all detail-free M and all queries T, there exists $\mathcal{I}$ such that $G = (M, \mathcal{I})$ has no strict truthful equilibrium.

How to get around this result?

Goal: overcome limitations of prior mechanisms.

How to get around this result?

Goal: overcome limitations of prior mechanisms.
Obstacle: Impossibility result!

How to get around this result?

Goal: overcome limitations of prior mechanisms.
Obstacle: Impossibility result!

Proposed solution: Output agreement mechanisms.

- simple to explain and implement

How to get around this result?

Goal: overcome limitations of prior mechanisms.
Obstacle: Impossibility result!

Proposed solution: Output agreement mechanisms.

- simple to explain and implement
- applicable in variety of complex domains

How to get around this result?

Goal: overcome limitations of prior mechanisms.
Obstacle: Impossibility result!

Proposed solution: Output agreement mechanisms.

- simple to explain and implement
- applicable in variety of complex domains
- detail-free

How to get around this result?

Goal: overcome limitations of prior mechanisms.
Obstacle: Impossibility result!

Proposed solution: Output agreement mechanisms.

- simple to explain and implement
- applicable in variety of complex domains
- detail-free
- unrestricted domain

How to get around this result?

Goal: overcome limitations of prior mechanisms.
Obstacle: Impossibility result!

Proposed solution: Output agreement mechanisms.

- simple to explain and implement
- applicable in variety of complex domains
- detail-free
- unrestricted domain
- ... but not truthful!

Outline

Information elicitation without verification

Formal setting and prior work

Impossibility results for IEWV

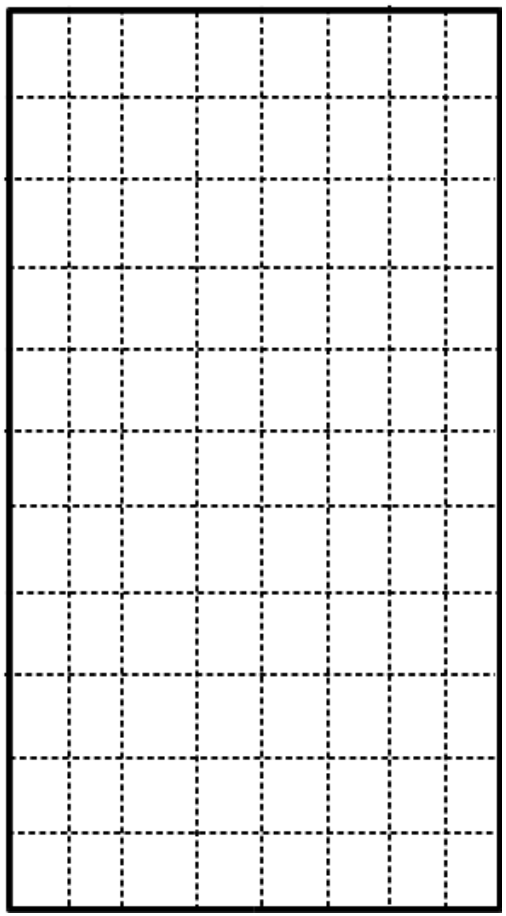
Output agreement mechanisms

Output agreement

Truthful $\rightarrow$ **common-knowledge truthful**:

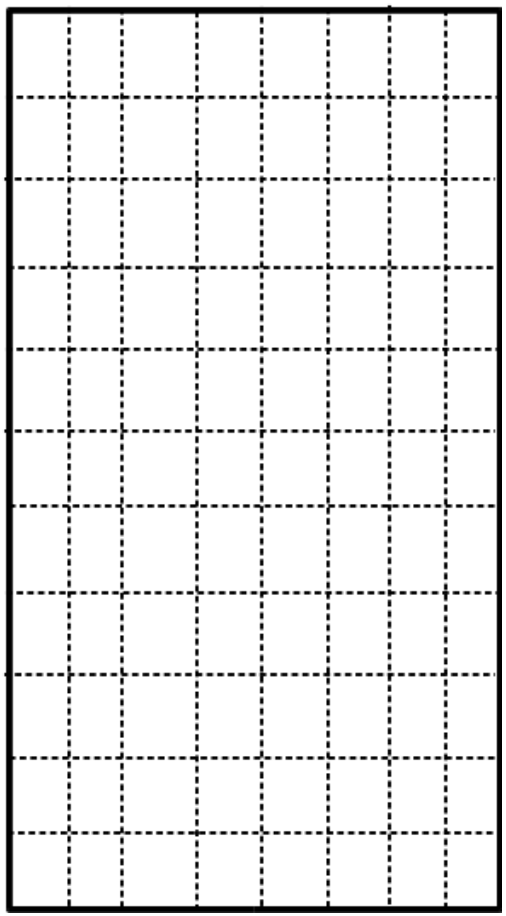
Common Knowledge

Ω: possible states of the world



Common Knowledge

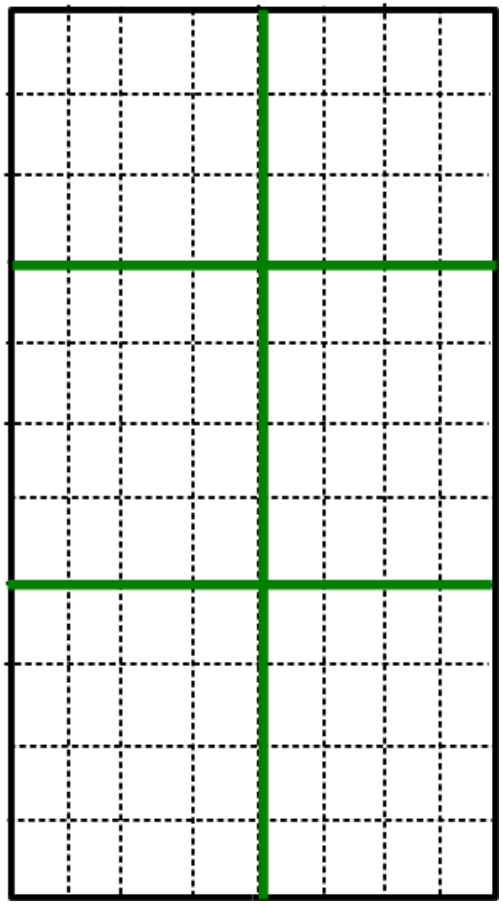
Ω: possible states of the world



$\mathcal{P}[\omega]$: common prior

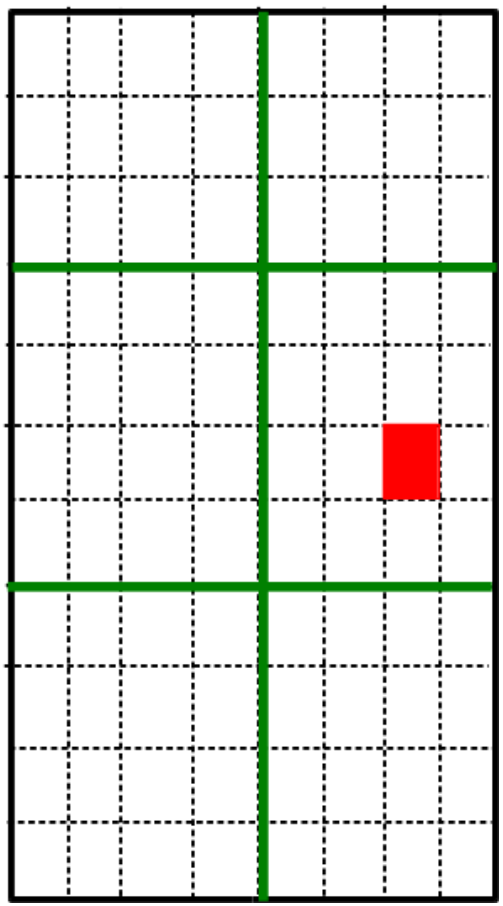
Common Knowledge

Π_1: player 1's partition



Common Knowledge

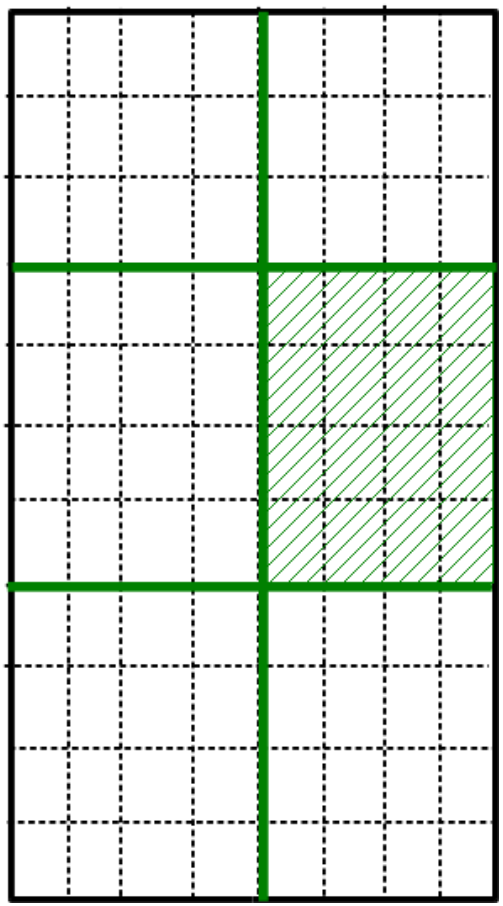
Π_1



ω^*: true state selected by nature

Common Knowledge

Π_1

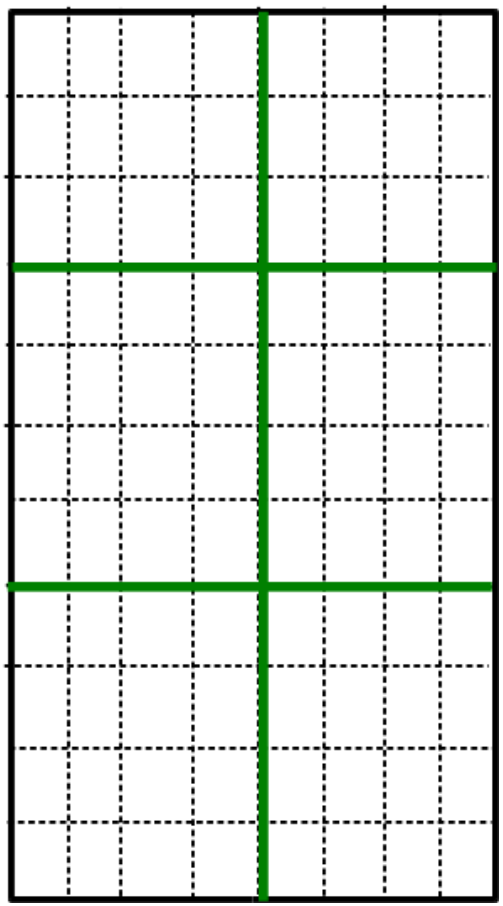


$\Pi_1(\omega^*)$: player 1's signal

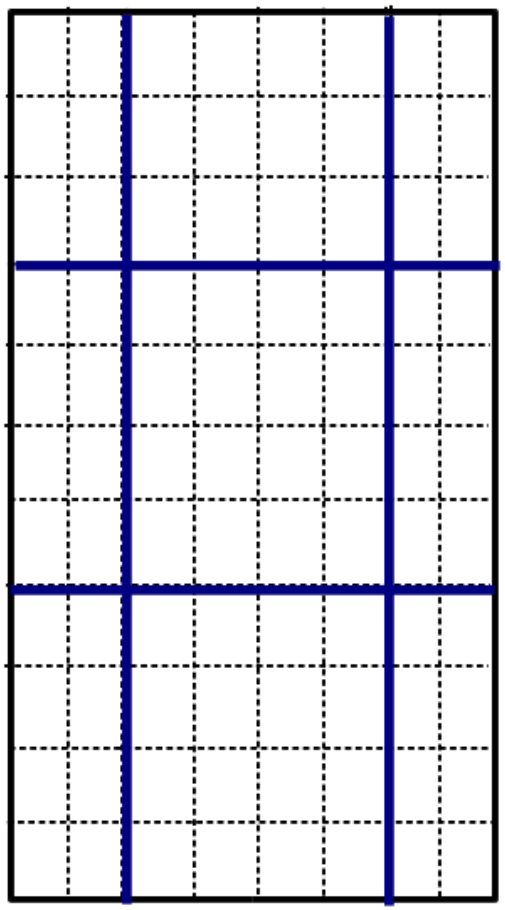
$\Pr[\omega \mid \Pi_1(\omega^*)]$: player 1's posterior

Common Knowledge

Π_1

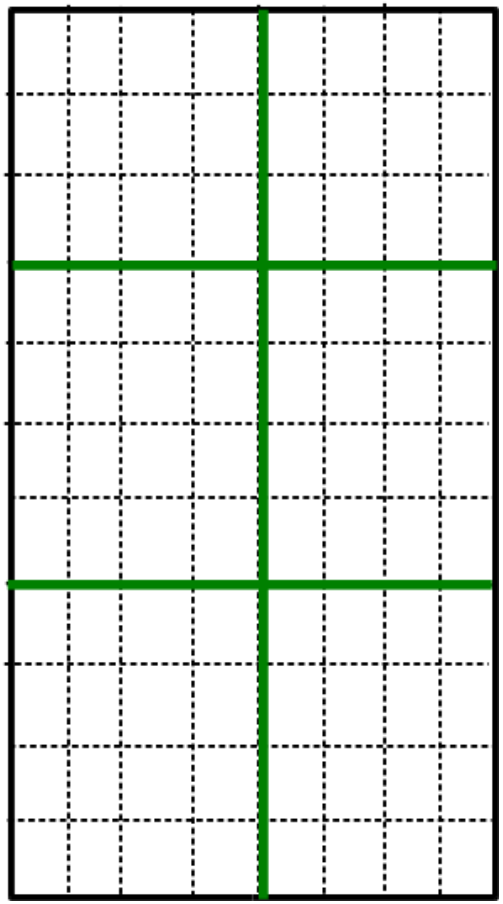


Π_2

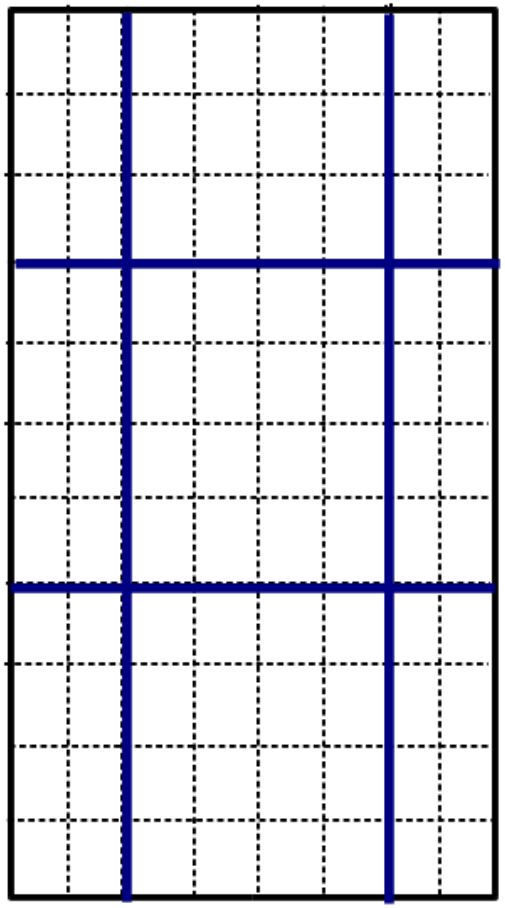


Common Knowledge

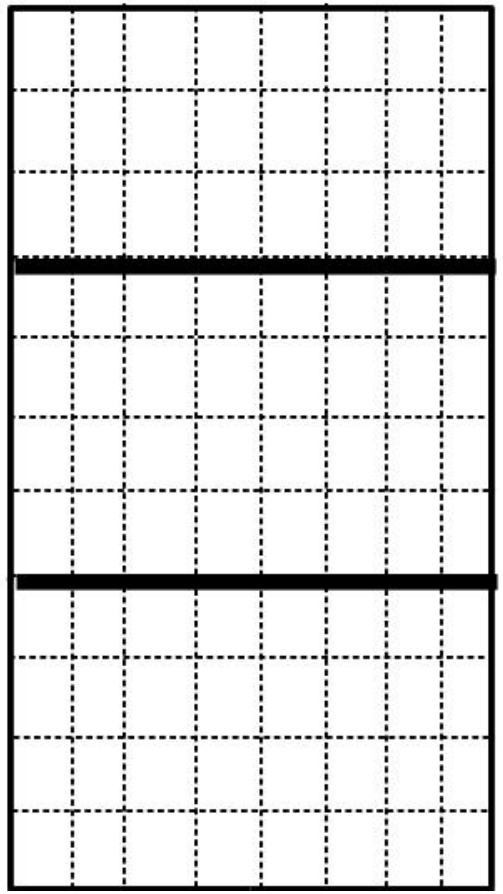
Π_1



Π_2



Π: common-knowledge partition



Output agreement

Truthful $\rightarrow$ **common-knowledge truthful**:
$s_i(\Pi_i(\omega^*)) = T(\Pi(\omega^*))$.
Previously: $= T(\Pi_i(\omega^*))$.

Output agreement: Origins

Output agreement: informally coined by von Ahn, Dabbish 2004.

Output agreement: Origins

Output agreement: informally coined by von Ahn, Dabbish 2004.

Game-theoretic analysis of ESP Game: Jain, Parkes 2008. (Specific agent model, not general output agreement framework.)

Output agreement: Origins

Output agreement: informally coined by von Ahn, Dabbish 2004.

Game-theoretic analysis of ESP Game: Jain, Parkes 2008. (Specific agent model, not general output agreement framework.)

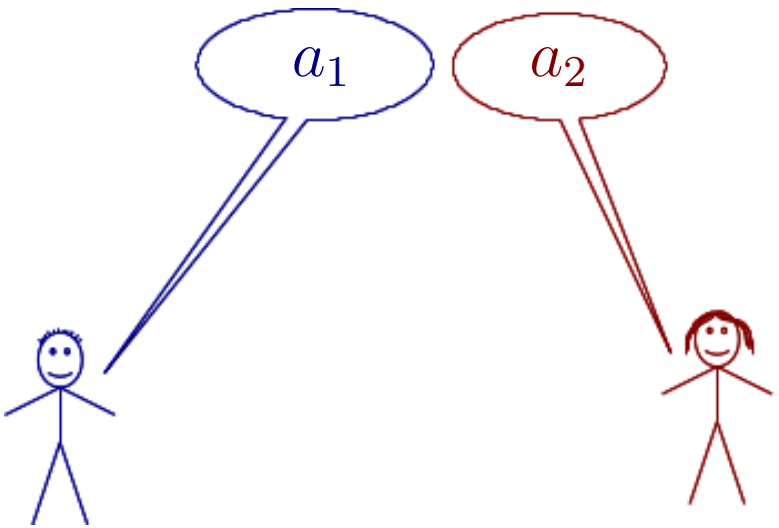
Here: first general formalization of output agreement.

Output agreement

An **output agreement mechanism**:

Output agreement

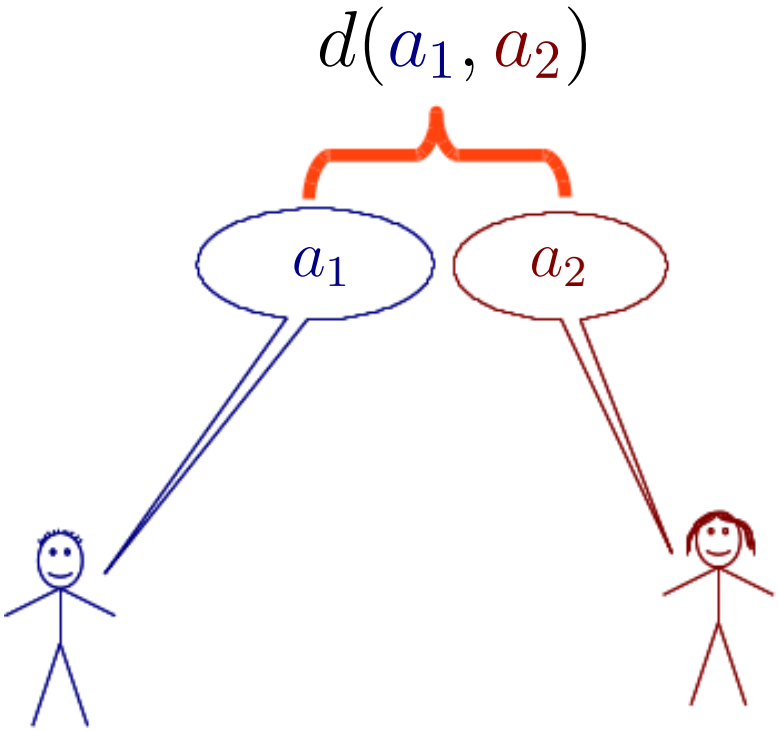
An **output agreement mechanism**:

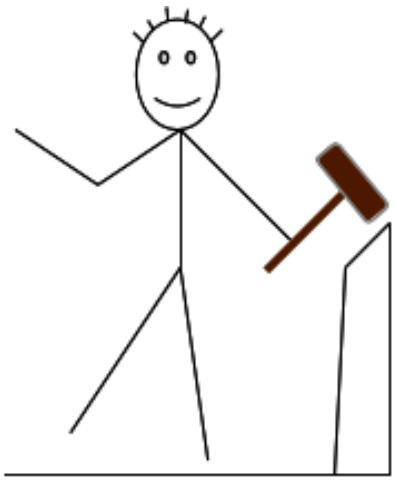


report space: A

Output agreement

An **output agreement mechanism**:

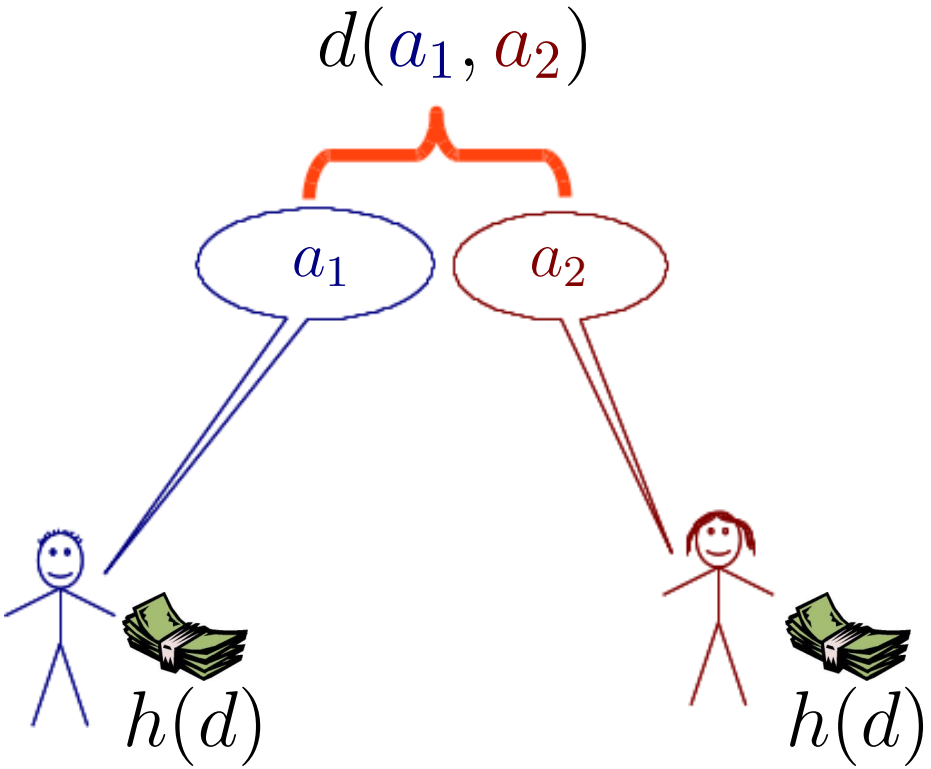


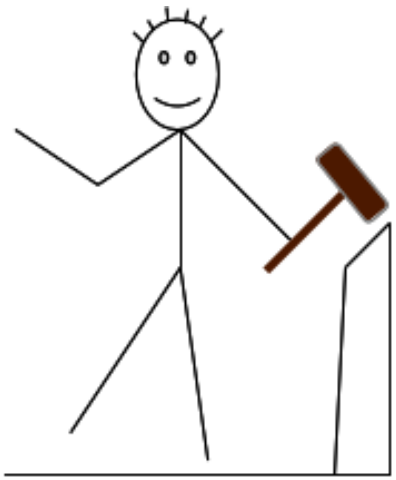


report space: (A, d)

Output agreement

An **output agreement mechanism**:





report space: (A, d)

payoff: h strictly decreasing

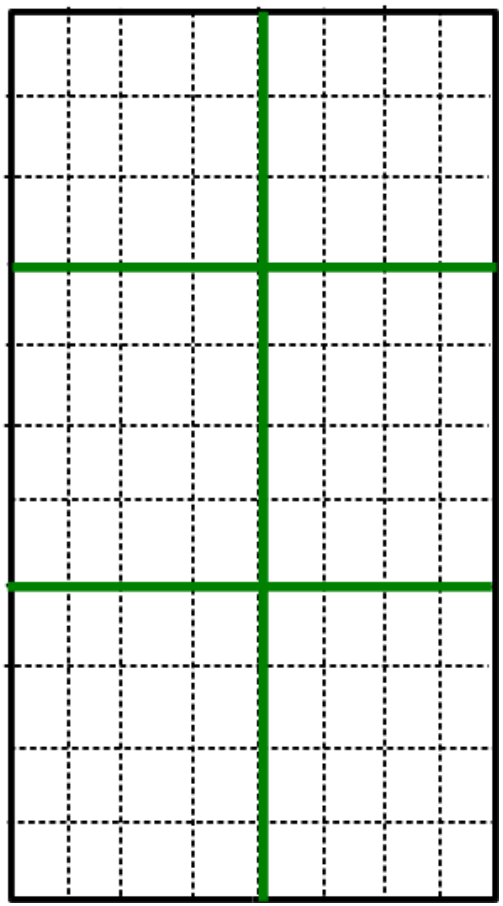
Output agreement

Theorem

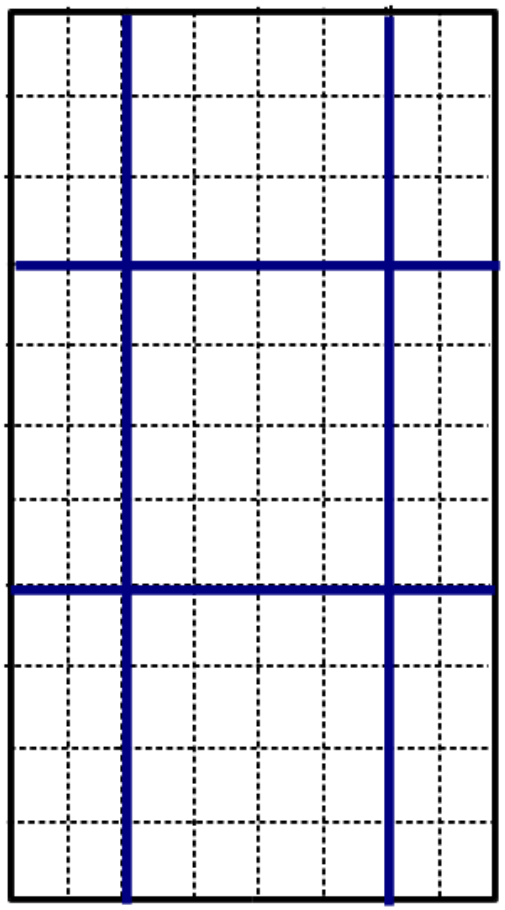
For any query T, there is an output agreement mechanism M eliciting a strict common-knowledge-truthful equilibrium.

Proof by picture

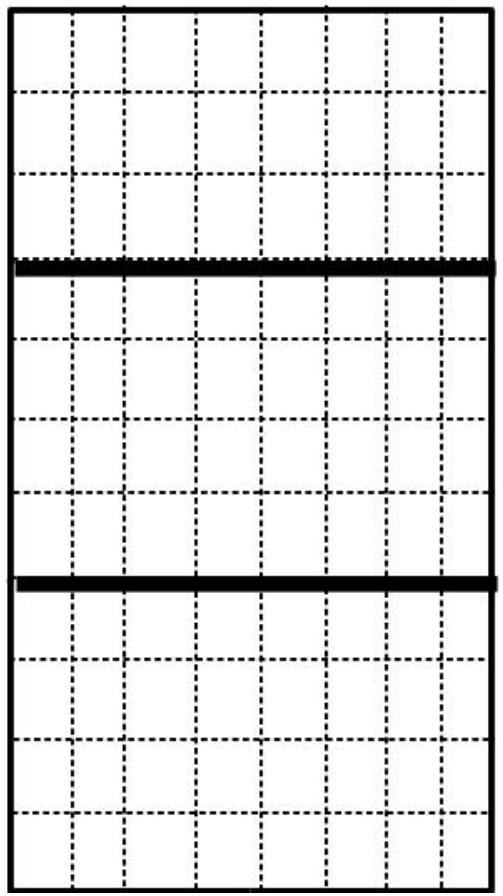
Π_1



Π_2



Π



Are "good" equilibria played?

What is "focal" in output agreement?

Are "good" equilibria played?

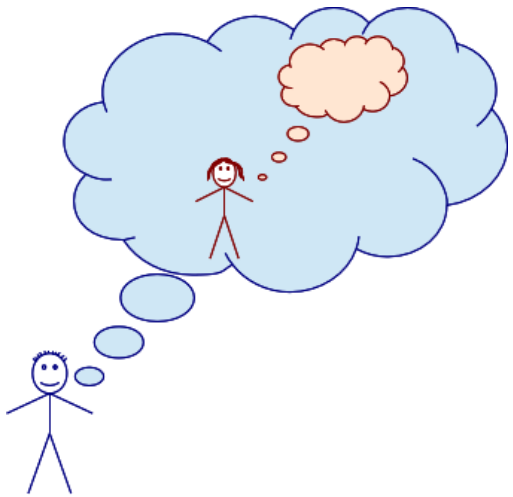
What is "focal" in output agreement?
One approach: *player inference*, beginning with truthful strategy.

Are "good" equilibria played?

What is "focal" in output agreement?
One approach: *player inference*, beginning with truthful strategy.

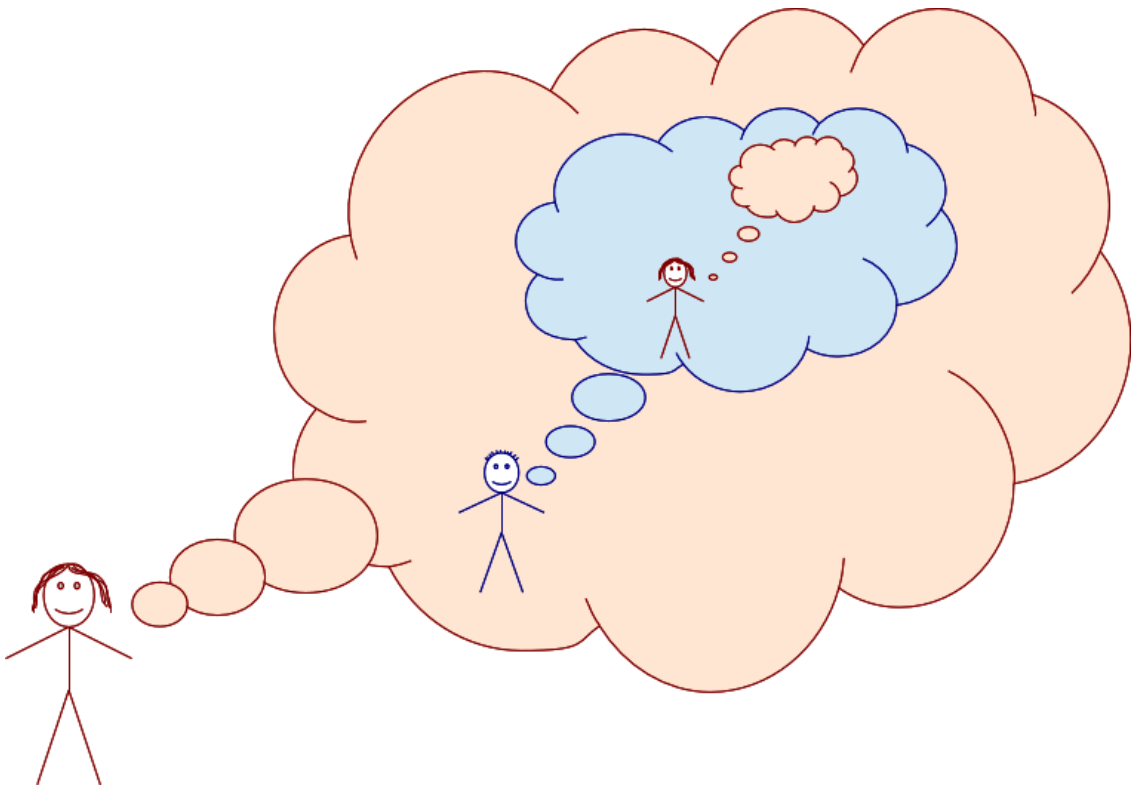
Are "good" equilibria played?

What is "focal" in output agreement?
One approach: *player inference*, beginning with truthful strategy.



Are "good" equilibria played?

What is "focal" in output agreement?
One approach: *player inference*, beginning with truthful strategy.



Output agreement

Inference: iteratively compute strategy that maximizes expected utility.

Output agreement

Inference: iteratively compute strategy that maximizes expected utility.

When does inference, starting with truthfulness, converge to common-knowledge truthfulness?

Output agreement

Inference: iteratively compute strategy that maximizes expected utility.

When does inference, starting with truthfulness, converge to common-knowledge truthfulness?

- Eliciting the mean: Yes!

Output agreement

Inference: iteratively compute strategy that maximizes expected utility.

When does inference, starting with truthfulness, converge to common-knowledge truthfulness?

- Eliciting the mean: Yes!
- Eliciting the median, mode: No!

Output agreement

Inference: iteratively compute strategy that maximizes expected utility.

When does inference, starting with truthfulness, converge to common-knowledge truthfulness?

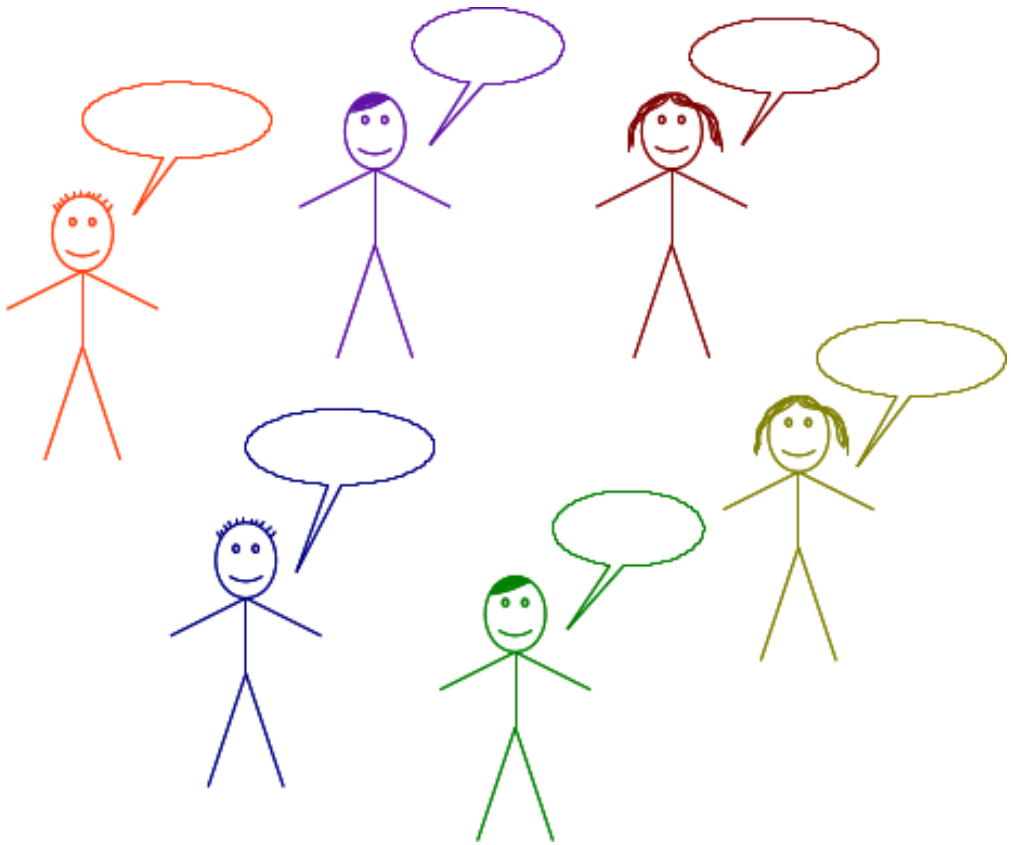
- Eliciting the mean: Yes!
- Eliciting the median, mode: No! (arbitrarily bad examples)

Output agreement

Mechanisms on many players?

Output agreement

Mechanisms on many players? (Yes)



Outline

Information elicitation without verification

Setting

Impossibility results

Output agreement

Summary

Summary

- IEWV: formalized mechanism design setting.

Summary

- IEWV: formalized mechanism design setting.
- (Almost) all mechanisms have bad equilibria.

Summary

- IEWV: formalized mechanism design setting.
- (Almost) all mechanisms have bad equilibria.
- There are no detail-free, unrestricted-domain, *truthful* mechanisms.

Summary

- IEWV: formalized mechanism design setting.
- (Almost) all mechanisms have bad equilibria.
- There are no detail-free, unrestricted-domain, *truthful* mechanisms.
- Output agreement:

Summary

- IEWV: formalized mechanism design setting.
- (Almost) all mechanisms have bad equilibria.
- There are no detail-free, unrestricted-domain, *truthful* mechanisms.
- Output agreement:
  - simple

Summary

- IEWV: formalized mechanism design setting.
- (Almost) all mechanisms have bad equilibria.
- There are no detail-free, unrestricted-domain, *truthful* mechanisms.
- Output agreement:
  - simple
  - applicable in complex domains

Summary

- IEWV: formalized mechanism design setting.
- (Almost) all mechanisms have bad equilibria.
- There are no detail-free, unrestricted-domain, *truthful* mechanisms.
- Output agreement:
  - simple
  - applicable in complex domains
  - detail-free, unrestricted-domain

Summary

- IEWV: formalized mechanism design setting.
- (Almost) all mechanisms have bad equilibria.
- There are no detail-free, unrestricted-domain, *truthful* mechanisms.
- Output agreement:
  - simple
  - applicable in complex domains
  - detail-free, unrestricted-domain
  - elicits *common knowledge*

Summary

- IEWV: formalized mechanism design setting.
- (Almost) all mechanisms have bad equilibria.
- There are no detail-free, unrestricted-domain, *truthful* mechanisms.
- Output agreement:
  - simple
  - applicable in complex domains
  - detail-free, unrestricted-domain
  - elicits *common knowledge*

Thanks!