Disagreement under almost common knowledge of rationality

Emiliano Catonini and Tatiana Mayskaya

HSE Moscow - ICEF

October 2019

Question

Aumann (1976) has shown that if two agents have common knowledge of

Question

Aumann (1976) has shown that if two agents have common knowledge of

- information structure
- rationality (Bayes rule)
- posteriors

Question

Aumann (1976) has shown that if two agents have common knowledge of

- information structure
- rationality (Bayes rule)
- posteriors

then their posteriors must be identical: they cannot "agree to disasgree".

Question

Aumann (1976) has shown that if two agents have common knowledge of

- information structure
- rationality (Bayes rule)
- posteriors

then their posteriors must be identical: they cannot "agree to disasgree".

But we do observe large disagreement among "intelligent" people.

Question

Aumann (1976) has shown that if two agents have common knowledge of

- information structure
- rationality (Bayes rule)
- posteriors

then their posteriors must be identical: they cannot "agree to disasgree".

But we do observe large disagreement among "intelligent" people.

Why?

Hypothesis

We retain common knowledge of the information structure

Hypothesis

We retain common knowledge of the information structure

We let agents share repeatedly not their posteriors, but a coarse opinion:

- Which of two states is more likely?

Hypothesis

We retain common knowledge of the information structure

We let agents share repeatedly not their posteriors, but a coarse opinion:

- Which of two states is more likely?
- Cave (1983), Bacharach (1985): this alone does not prevent agreement (on which state is more likely)

Hypothesis

We retain common knowledge of the information structure

We let agents share repeatedly not their posteriors, but a coarse opinion:

- Which of two states is more likely?
- Cave (1983), Bacharach (1985): this alone does not prevent agreement (on which state is more likely)

We consider a small deviation from common knowledge of rationality.

Conclusion

That this small deviation may induce a small disagreement is not surprising.

But no matter how small the deviation is, very large disagreements persist!

Conclusion

That this small deviation may induce a small disagreement is not surprising.

But no matter how small the deviation is, very large disagreements persist!

Result in the spirit of Rubinstein's (1989) Email Game: small deviation from common knowledge of an event causes a big departure in the predictions.

Explanation

As time goes and disagreement persists, agents with strong opinions question the cognitive ability/the order of belief in rationality of the opponent.

Explanation

As time goes and disagreement persists, agents with strong opinions question the cognitive ability/the order of belief in rationality of the opponent.

The initial deviation from common belief in rationality becomes larger!

Explanation

As time goes and disagreement persists, agents with strong opinions question the cognitive ability/the order of belief in rationality of the opponent.

The initial deviation from common belief in rationality becomes larger!

With belief updating about the state *but not* about rationality, the size of disagreement would be bounded above by the size of the initial doubt

Game

There is a binary state of the world $\theta \in \{0, 1\}$ and two agents $i = 1, 2$.

Game

There is a binary state of the world $\theta \in \{0, 1\}$ and two agents $i = 1, 2$.

At time $t = 0$, the agents receive a private signal about the state.

Game

There is a binary state of the world $\theta \in \{0, 1\}$ and two agents $i = 1, 2$.

At time $t = 0$, the agents receive a private signal about the state.

At each time $t > 0$, the agents simultaneously make a statement: 0 or 1.

- Past statements are perfectly observed.

Game

There is a binary state of the world $\theta \in \{0, 1\}$ and two agents $i = 1, 2$.

At time $t = 0$, the agents receive a private signal about the state.

At each time $t > 0$, the agents simultaneously make a statement: 0 or 1.

- Past statements are perfectly observed.

We assume that in each period the agents have the incentive to guess the state they consider more likely.

Beliefs about rationality

Rationality = Bayes rule

We focus on agents who believe that, with probability ε, the opponent is a *level-k thinker* (Stahl and Wilson 1995)

Beliefs about rationality

Rationality = Bayes rule

We focus on agents who believe that, with probability ε, the opponent is a *level-k thinker* (Stahl and Wilson 1995)

Level 1 is rational but does not believe/fails to see that also the opponent is.

- Alternatively, level 1 could be a committed type who insists on 0 or 1 for some reason.

Beliefs about rationality

Rationality = Bayes rule

We focus on agents who believe that, with probability ε, the opponent is a *level-k thinker* (Stahl and Wilson 1995)

Level 1 is rational but does not believe/fails to see that also the opponent is.

- Alternatively, level 1 could be a committed type who insists on 0 or 1 for some reason.

Level k uses Bayes rule and believes that the opponent is level k-1.

Beliefs about rationality

Rationality = Bayes rule

We focus on agents who believe that, with probability ε, the opponent is a *level-k thinker* (Stahl and Wilson 1995)

Level 1 is rational but does not believe/fails to see that also the opponent is.

- Alternatively, level 1 could be a committed type who insists on 0 or 1 for some reason.

Level k uses Bayes rule and believes that the opponent is level k-1.

For a level infinite agent there is common belief in rationality.

Beliefs about rationality

Rationality = Bayes rule

We focus on agents who believe that, with probability ε, the opponent is a *level-k thinker* (Stahl and Wilson 1995)

Level 1 is rational but does not believe/fails to see that also the opponent is.

- Alternatively, level 1 could be a committed type who insists on 0 or 1 for some reason.

Level k uses Bayes rule and believes that the opponent is level k-1.

For a level infinite agent there is common belief in rationality.

To fix ideas, we assign probability $1-\varepsilon$ to the opponent being level infinite.

- "Smallest" departure from common knowledge, but it does not really matter where $1-\varepsilon$ goes.

Beliefs about rationality

Rationality = Bayes rule

We focus on agents who believe that, with probability ε, the opponent is a *level-k thinker* (Stahl and Wilson 1995)

Level 1 is rational but does not believe/fails to see that also the opponent is.

- Alternatively, level 1 could be a committed type who insists on 0 or 1 for some reason.

Level k uses Bayes rule and believes that the opponent is level k-1.

For a level infinite agent there is common belief in rationality.

To fix ideas, we assign probability $1-\varepsilon$ to the opponent being level infinite.

- "Smallest" departure from common knowledge, but it does not really matter where $1-\varepsilon$ goes.
- Alternatives: cognitive hieararchical model (Camerer et al. 2004), common-p belief of rationality (Monderer&Samet 1989, Hu 2007).

Result

Theorem

For any non-degenerate prior p about θ, any level $k \geq 1$, any positive amount of doubt $\varepsilon \in (0, 1)$, and any final level of disagreement $d \in (0, 1)$, there exist a signal structure and realizations of the signals s^0 and s^1 such that our agents permanently disagree with each other, and the distance between their posteriors exceeds d forever.

Proof

Each agent receives $n \geq 2^{k+1}$ binary signals, each correct with probability $q > p$.

Proof

Each agent receives $n \geq 2^{k+1}$ binary signals, each correct with probability $q > p$.

A level-1 will permanently say 0 if less than 1/2 of her signals are ones;

Proof

Each agent receives $n \geq 2^{k+1}$ binary signals, each correct with probability $q > p$.

A level-1 will permanently say 0 if less than 1/2 of her signals are ones;
A level-2 will permanently say 1 if less than 1/4 of her signals are zeros;

Proof

Each agent receives $n \geq 2^{k+1}$ binary signals, each correct with probability $q > p$.

A level-1 will permanently say 0 if less than $1/2$ of her signals are ones;
A level-2 will permanently say 1 if less than $1/4$ of her signals are zeros;
A level-3 will permanently say 0 if less than $1/8$ of her signals are ones;

Proof

Each agent receives $n \geq 2^{k+1}$ binary signals, each correct with probability $q > p$.

A level-1 will permanently say 0 if less than $1/2$ of her signals are ones;
A level-2 will permanently say 1 if less than $1/4$ of her signals are zeros;
A level-3 will permanently say 0 if less than $1/8$ of her signals are ones;
...
A level-k will permanently say 0 if less than $n/2^k \geq 2$ signals are ones.

Now focus on our "smart but doubtful" agents.

Now focus on our "smart but doubtful" agents.

Let s^i denote the number of signals 1 received by agent i.

Focus on agents with $s^0 = 0$ and $s^1 = n$, and take the viewpoint of agent 1.

Now focus on our "smart but doubtful" agents.

Let s^i denote the number of signals 1 received by agent i.

Focus on agents with $s^0 = 0$ and $s^1 = n$, and take the viewpoint of agent 1.

Wlog, assume that state 0 is weakly more likely a priori.

We show that for every $q > 1/2$ and some $n \geq 2^{k+1}$, at every history of disagreement $h^t = (0, 1)^t$, agent 1 assigns probability higher than $(1 + d)/2$ to state 1.

Now focus on our "smart but doubtful" agents.

Let s^i denote the number of signals 1 received by agent i.

Focus on agents with $s^0 = 0$ and $s^1 = n$, and take the viewpoint of agent 1.

Wlog, assume that state 0 is weakly more likely a priori.

We show that for every $q > 1/2$ and some $n \geq 2^{k+1}$, at every history of disagreement $h^t = (0, 1)^t$, agent 1 assigns probability higher than $(1 + d)/2$ to state 1.

Then, the difference in posteriors remains perpetually higher than

$$(1 + d)/2 - (1 - (1 + d)/2) = d.$$

$$\Pr\left[\theta = 1 \mid h_t,\ s^1 = n\right]$$

$$
\begin{aligned}
& \Pr\left[\theta = 1 \mid h_t,\ s^1 = n\right] \\
= \quad & \sum_{m=0}^{n} \Pr\left[\theta = 1 \mid s^1 = n, s^0 = m\right] \Pr\left[s^0 = m \mid h_t,\ s^1 = n\right]
\end{aligned}
$$

$$
\begin{aligned}
& \Pr\left[\theta = 1 \mid h_t,\ s^1 = n\right] \\
= \quad & \sum_{m=0}^{n} \Pr\left[\theta = 1 \mid s^1 = n, s^0 = m\right] \Pr\left[s^0 = m \mid h_t,\ s^1 = n\right] \\
\geq \quad & \sum_{m=0,1} \frac{\Pr\left[\theta = 1 \mid s^1 = n, s^0 = m\right] \Pr\left[s^0 = m \mid h_t,\ s^1 = n\right]}{\Pr\left[s^0 = 0 \mid h_t,\ s^1 = n\right] + \Pr\left[s^0 = 1 \mid h_t,\ s^1 = n\right]}
\end{aligned}
$$

$$
\begin{aligned}
& \Pr\left[\theta = 1 \mid h_t,\ s^1 = n\right] \\
= & \sum_{m=0}^{n} \Pr\left[\theta = 1 \mid s^1 = n, s^0 = m\right] \Pr\left[s^0 = m \mid h_t,\ s^1 = n\right] \\
\geq & \sum_{m=0,1} \frac{\Pr\left[\theta = 1 \mid s^1 = n, s^0 = m\right] \Pr\left[s^0 = m \mid h_t,\ s^1 = n\right]}{\Pr\left[s^0 = 0 \mid h_t,\ s^1 = n\right] + \Pr\left[s^0 = 1 \mid h_t,\ s^1 = n\right]} \\
\geq & \frac{\Pr\left[\theta = 1 \mid {s^1=n \atop s^0=0}\right] \Pr\left[s^0 = 0 \mid {h_t, \atop s^1=n}\right] + \Pr\left[\theta = 1 \mid {s^1=n \atop s^0=1}\right] \Pr\left[{s^0=1 \atop L_0=k} \mid {h_t, \atop s^1=n}\right]}{\Pr\left[s^0 = 0 \mid h_t,\ s^1 = n\right] + \Pr\left[s^0 = 1,\ L_0 = k \mid h_t,\ s^1 = n\right]}
\end{aligned}
$$

$$
\begin{aligned}
& \Pr\left[\theta = 1 \mid h_t,\ s^1 = n\right] \\
= \quad & \sum_{m=0}^{n} \Pr\left[\theta = 1 \mid s^1 = n, s^0 = m\right] \Pr\left[s^0 = m \mid h_t,\ s^1 = n\right] \\
\geq \quad & \sum_{m=0,1} \frac{\Pr\left[\theta = 1 \mid s^1 = n, s^0 = m\right] \Pr\left[s^0 = m \mid h_t,\ s^1 = n\right]}{\Pr\left[s^0 = 0 \mid h_t,\ s^1 = n\right] + \Pr\left[s^0 = 1 \mid h_t,\ s^1 = n\right]} \\
\geq \quad & \frac{\Pr\left[\theta = 1 \mid {s^1=n \atop s^0=0}\right] \Pr\left[s^0 = 0 \mid {h_t, \atop s^1=n}\right] + \Pr\left[\theta = 1 \mid {s^1=n \atop s^0=1}\right] \Pr\left[{s^0=1 \atop L_0=k} \mid {h_t, \atop s^1=n}\right]}{\Pr\left[s^0 = 0 \mid h_t,\ s^1 = n\right] + \Pr\left[s^0 = 1,\ L_0 = k \mid h_t,\ s^1 = n\right]} \\
\geq \quad & \frac{\Pr\left[\theta = 1 \mid {s^1=n \atop s^0=0}\right] \Pr\left[s^0 = 0 \mid s^1 = n\right] + \Pr\left[\theta = 1 \mid {s^1=n \atop s^0=1}\right] \Pr\left[{s^0=1 \atop L_0=k} \mid s^1 = n\right]}{\Pr\left[s^0 = 0 \mid s^1 = n\right] + \Pr\left[s^0 = 1,\ L_0 = k \mid s^1 = n\right]}
\end{aligned}
$$

$$
\begin{aligned}
& \Pr\left[\theta = 1 \mid h_t,\ s^1 = n\right] \\
= \quad & \sum_{m=0}^{n} \Pr\left[\theta = 1 \mid s^1 = n, s^0 = m\right] \Pr\left[s^0 = m \mid h_t,\ s^1 = n\right] \\
\geq \quad & \sum_{m=0,1} \frac{\Pr\left[\theta = 1 \mid s^1 = n, s^0 = m\right] \Pr\left[s^0 = m \mid h_t,\ s^1 = n\right]}{\Pr\left[s^0 = 0 \mid h_t,\ s^1 = n\right] + \Pr\left[s^0 = 1 \mid h_t,\ s^1 = n\right]} \\
\geq \quad & \frac{\Pr\left[\theta = 1 \mid {s^1=n \atop s^0=0}\right] \Pr\left[s^0 = 0 \mid {h_t, \atop s^1=n}\right] + \Pr\left[\theta = 1 \mid {s^1=n \atop s^0=1}\right] \Pr\left[{s^0=1 \atop L_0=k} \mid {h_t, \atop s^1=n}\right]}{\Pr\left[s^0 = 0 \mid h_t,\ s^1 = n\right] + \Pr\left[s^0 = 1,\ L_0 = k \mid h_t,\ s^1 = n\right]} \\
\geq \quad & \frac{\Pr\left[\theta = 1 \mid {s^1=n \atop s^0=0}\right] \Pr\left[s^0 = 0 \mid s^1 = n\right] + \Pr\left[\theta = 1 \mid {s^1=n \atop s^0=1}\right] \Pr\left[{s^0=1 \atop L_0=k} \mid s^1 = n\right]}{\Pr\left[s^0 = 0 \mid s^1 = n\right] + \Pr\left[s^0 = 1,\ L_0 = k \mid s^1 = n\right]} \\
= \quad & \frac{\Pr\left[\theta = 1,\ s^1 = n,\ s^0 = 0\right] + \Pr\left[\theta = 1,\ s^1 = n,\ s^0 = 1,\ L_0 = k\right]}{\Pr\left[s^0 = 0,\ s^1 = n\right] + \Pr\left[s^0 = 1,\ L_0 = k,\ s^1 = n\right]}
\end{aligned}
$$

$$
\begin{aligned}
& \Pr\left[\theta = 1 \mid h_t,\ s^1 = n\right] \\
= & \sum_{m=0}^{n} \Pr\left[\theta = 1 \mid s^1 = n, s^0 = m\right] \Pr\left[s^0 = m \mid h_t,\ s^1 = n\right] \\
\geq & \sum_{m=0,1} \frac{\Pr\left[\theta = 1 \mid s^1 = n, s^0 = m\right] \Pr\left[s^0 = m \mid h_t,\ s^1 = n\right]}{\Pr\left[s^0 = 0 \mid h_t,\ s^1 = n\right] + \Pr\left[s^0 = 1 \mid h_t,\ s^1 = n\right]} \\
\geq & \frac{\Pr\left[\theta = 1 \mid {s^1=n \atop s^0=0}\right] \Pr\left[s^0 = 0 \mid {h_t, \atop s^1=n}\right] + \Pr\left[\theta = 1 \mid {s^1=n \atop s^0=1}\right] \Pr\left[{s^0=1 \atop L_0=k} \mid {h_t, \atop s^1=n}\right]}{\Pr\left[s^0 = 0 \mid h_t,\ s^1 = n\right] + \Pr\left[s^0 = 1,\ L_0 = k \mid h_t,\ s^1 = n\right]} \\
\geq & \frac{\Pr\left[\theta = 1 \mid {s^1=n \atop s^0=0}\right] \Pr\left[s^0 = 0 \mid s^1 = n\right] + \Pr\left[\theta = 1 \mid {s^1=n \atop s^0=1}\right] \Pr\left[{s^0=1 \atop L_0=k} \mid s^1 = n\right]}{\Pr\left[s^0 = 0 \mid s^1 = n\right] + \Pr\left[s^0 = 1,\ L_0 = k \mid s^1 = n\right]} \\
= & \frac{\Pr\left[\theta = 1,\ s^1 = n,\ s^0 = 0\right] + \Pr\left[\theta = 1,\ s^1 = n,\ s^0 = 1,\ L_0 = k\right]}{\Pr\left[s^0 = 0,\ s^1 = n\right] + \Pr\left[s^0 = 1,\ L_0 = k,\ s^1 = n\right]} \\
= & \frac{\left(1 + \varepsilon n \frac{q}{1-q}\right) \Pr\left[\theta = 1\right]}{1 + \varepsilon n \left(\frac{q}{1-q} \Pr\left[\theta = 1\right] + \frac{1-q}{q} \Pr\left[\theta = 0\right]\right)} \longrightarrow 1.
\end{aligned}
$$

On the information structure

As we raise k or d, or as we lower ε, we need higher n — the game "changes".

On the information structure

As we raise k or d, or as we lower ε, we need higher n — the game "changes".

But it would be equivalent to have one continuous signal $\sigma_i \in [0, 1]$ for each agent i, where $\sigma_i = 0, 1$ is evidence of θ.

Then, as we raise k or d, or as we lower ε, we can just focus on agents who receive signals closer and closer to 0 and 1.

On strategic considerations

Consider patient agents who care about the stream of correct guesses.

On strategic considerations

Consider patient agents who care about the stream of correct guesses.

Suppose to be on the permanent disagreement path: no surprise has taken place.

On strategic considerations

Consider patient agents who care about the stream of correct guesses.

Suppose to be on the permanent disagreement path: no surprise has taken place.

Level-k agents with sufficiently strong evidence of a state find it impossible to elicit any useful information from the opponent.

- They think the opponent is more stubborn, therefore will never reveal stronger opposite evidence.

On strategic considerations

Consider patient agents who care about the stream of correct guesses.

Suppose to be on the permanent disagreement path: no surprise has taken place.

Level-k agents with sufficiently strong evidence of a state find it impossible to elicit any useful information from the opponent.

- They think the opponent is more stubborn, therefore will never reveal stronger opposite evidence.

Our "smart but doubtful" agents have no incentive to elicit with all zeros.

On strategic considerations

Consider patient agents who care about the stream of correct guesses.

Suppose to be on the permanent disagreement path: no surprise has taken place.

Level-k agents with sufficiently strong evidence of a state find it impossible to elicit any useful information from the opponent.

- They think the opponent is more stubborn, therefore will never reveal stronger opposite evidence.

Our “smart but doubtful” agents have no incentive to elicit with all zeros.

With all ones, they would like to know if the opponent has all zeros or not.

On strategic considerations

Consider patient agents who care about the stream of correct guesses.

Suppose to be on the permanent disagreement path: no surprise has taken place.

Level-k agents with sufficiently strong evidence of a state find it impossible to elicit any useful information from the opponent.

- They think the opponent is more stubborn, therefore will never reveal stronger opposite evidence.

Our "smart but doubtful" agents have no incentive to elicit with all zeros.

With all ones, they would like to know if the opponent has all zeros or not.

Can lying help to achieve this?