Disagreement Under Almost Common Knowledge of Rationality

Emiliano Catonini   Tatiana Mayskaya

Higher School of Economics

19th Annual SAET Conference
1 July 2019

Paradox

Aumann (1976)

If two individuals share the same prior and have common knowledge among them of

- their information partitions
- their rationality (that is, both agents use Bayes rule to update their beliefs)
- their posteriors

then these posteriors must be the same (they cannot agree to disagree)

Question : Why do we disagree?

Our answer : People disagree because they doubt each other's rationality (and they doubt each other's rationality because they disagree)

Paradox

Aumann (1976)

If two individuals share the same prior and have common knowledge among them of

- their information partitions
- their rationality (that is, both agents use Bayes rule to update their beliefs)
- their posteriors

then these posteriors must be the same (they cannot agree to disagree)

Paradox

Aumann (1976)

If two individuals share the same prior and have common knowledge among them of

- their information partitions
- their rationality (that is, both agents use Bayes rule to update their beliefs)
- their posteriors

then these posteriors must be the same (they cannot agree to disagree)

Paradox

Aumann (1976)

If two individuals share the same prior and have common knowledge among them of

- their information partitions
- their rationality (that is, both agents use Bayes rule to update their beliefs)
- their posteriors

then these posteriors must be the same (they cannot agree to disagree)

Paradox

Aumann (1976)

If two individuals share the same prior and have common knowledge among them of

- their information partitions
- their rationality (that is, both agents use Bayes rule to update their beliefs)
- their posteriors

then these posteriors must be the same (they cannot agree to disagree)

Paradox

Aumann (1976)

If two individuals share the same prior and have common knowledge among them of

- their information partitions
- their rationality (that is, both agents use Bayes rule to update their beliefs)
- their posteriors

then these posteriors must be the same (they cannot agree to disagree)

Paradox

Aumann (1976)

If two individuals share the same prior and have common knowledge among them of

- their information partitions
- their rationality (that is, both agents use Bayes rule to update their beliefs)
- their posteriors

then these posteriors must be the same (they cannot agree to disagree)

Robustness:

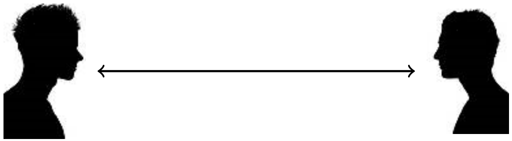
- Geanakoplos and Polemarchakis (1982): c.k. of rationality + communication of posteriors
- Cave (1983), Bacharach (1985): c.k. of rationality + communication of decisions among like-minded agents

Result (informally)

People disagree because they stop “hearing” each other, stop updating their beliefs. That happens if we relax the assumption of **common knowledge** of rationality.

Result (informally)

People disagree because they stop "hearing" each other, stop updating their beliefs. That happens if we relax the assumption of **common knowledge** of rationality.



Result (informally)

People disagree because they stop "hearing" each other, stop updating their beliefs. That happens if we relax the assumption of **common knowledge** of rationality.

Result (informally)

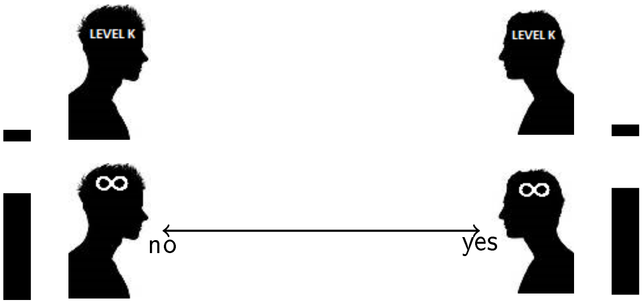
People disagree because they stop “hearing” each other, stop updating their beliefs. That happens if we relax the assumption of **common knowledge** of rationality.

Result (informally)

People disagree because they stop “hearing” each other, stop updating their beliefs. That happens if we relax the assumption of **common knowledge** of rationality.

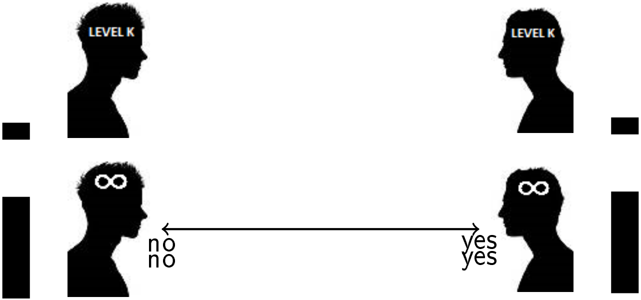
Result (informally)

People disagree because they stop "hearing" each other, stop updating their beliefs. That happens if we relax the assumption of **common knowledge** of rationality.



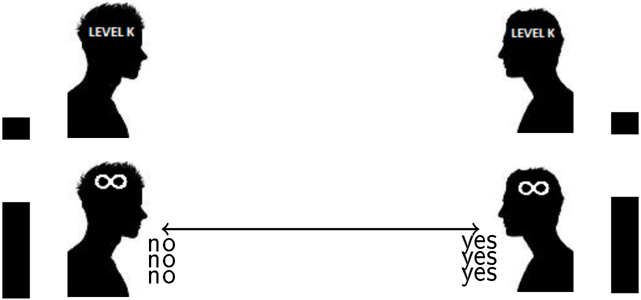
Result (informally)

People disagree because they stop "hearing" each other, stop updating their beliefs. That happens if we relax the assumption of **common knowledge** of rationality.



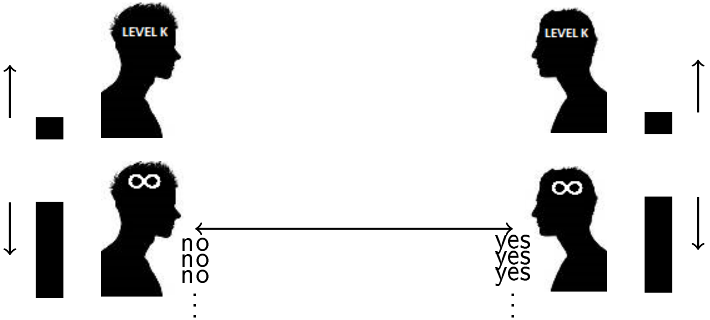
Result (informally)

People disagree because they stop “hearing” each other, stop updating their beliefs. That happens if we relax the assumption of **common knowledge** of rationality.



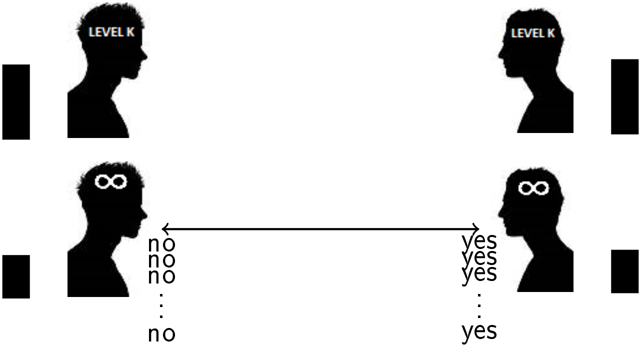
Result (informally)

People disagree because they stop "hearing" each other, stop updating their beliefs. That happens if we relax the assumption of **common knowledge** of rationality.



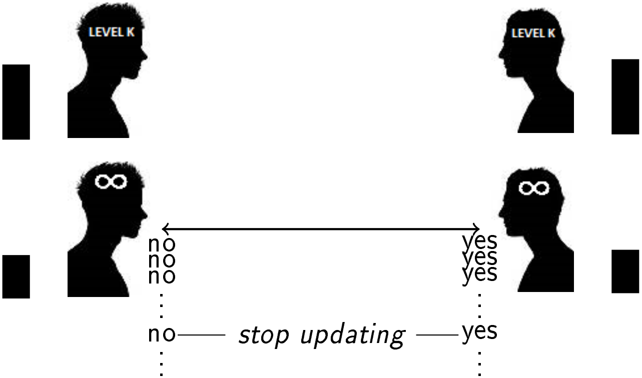
Result (informally)

People disagree because they stop "hearing" each other, stop updating their beliefs. That happens if we relax the assumption of **common knowledge** of rationality.



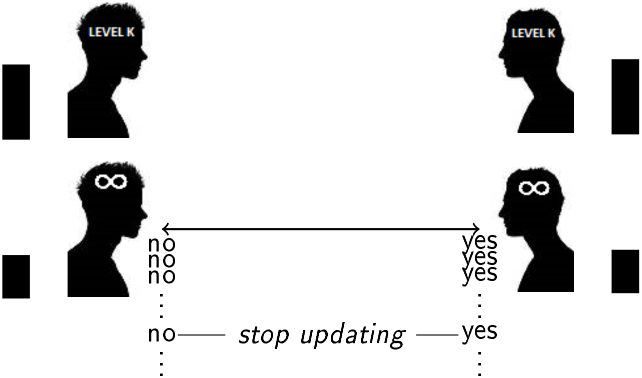
Result (informally)

People disagree because they stop “hearing” each other, stop updating their beliefs. That happens if we relax the assumption of **common knowledge** of rationality.



Result (informally)

People disagree because they stop "hearing" each other, stop updating their beliefs. That happens if we relax the assumption of **common knowledge** of rationality.



People disagree because they doubt each other's rationality (and they doubt each other's rationality because they disagree)

Model

Communication game

- State $\theta \in \{0, 1\}$

Model

Communication game

- State $\theta \in \{0, 1\}$
- Two agents exchange their guesses (0 or 1) about the state each period $t = 1, 2, \ldots$

Model

Communication game

- State $\theta \in \{0, 1\}$
- Two agents exchange their guesses (0 or 1) about the state each period $t = 1, 2, \ldots$
  - assume the agents guess the state sincerely

- State $\theta \in \{0, 1\}$
- Two agents exchange their guesses (0 or 1) about the state each period $t = 1, 2, \ldots$
  - assume the agents guess the state sincerely
- Period 1: common prior + private signals

Model

Belief formation

Based on the cognitive hierarchical (CH) theory proposed in Camerer et al (2004)

Level 0 agent randomizes each period: $\begin{cases} 0, & \text{with probability } 1/2 \\ 1, & \text{with probability } 1/2 \end{cases}$

Level 1 agent thinks that his opponent is of level 0

Model

Belief formation

Based on the cognitive hierarchical (CH) theory proposed in Camerer et al (2004)

Level 0 agent randomizes each period: $\begin{cases} 0, & \text{with probability } 1/2 \\ 1, & \text{with probability } 1/2 \end{cases}$

Level 1 agent thinks that his opponent is of level 0

Parameter of the model: function f — probability distribution over the levels of thinking

$$f(k) \in [0,1], \quad \sum_{k=0}^{+\infty} f(k) = 1$$

Level $k \geq 2$ agent thinks that the level of thinking of his opponent is distributed over 0 through $k-1$ according to the distribution f truncated at level $k-1$.

Model

Belief formation

Based on the cognitive hierarchical (CH) theory proposed in Camerer et al (2004)

Level 0 agent randomizes each period: $\begin{cases} 0, & \text{with probability } 1/2 \\ 1, & \text{with probability } 1/2 \end{cases}$

Level 1 agent thinks that his opponent is of level 0

Parameter of the model: function f — probability distribution over the levels of thinking

$$f(k) \in [0,1], \quad \sum_{k=0}^{+\infty} f(k) = 1$$

Level $k \geq 2$ agent thinks that the level of thinking of his opponent is distributed over 0 through $k-1$ according to the distribution f truncated at level $k-1$.

Novelty: f might be inconsistent with realized distribution. We assume both agents are of level ∞. Alternative interpretation: we make predictions only for ∞-level agents.

Result (formally)

Theorem

For any level $k \geq 1$, any positive amount of doubt $\varepsilon \in (0, 1)$, any "correct" posterior $\rho = \mathbb{P}\left[\theta = 1 \mid s^{(1)}, s^{(2)}\right] \in (0, 0.5)$, we can find a prior, a signal structure $F_i^{(\theta)}$, $i \in \{1, 2\}$, $\theta \in \{0, 1\}$, and the realization of the signals $s^{(1)}$ and $s^{(2)}$ such that the ∞-level agents with $f(0) = \ldots = f(k-1) = 0$ and $f(k) = \varepsilon$ permanently disagree with each other.

Theorem

For any level $k \geq 1$, any positive amount of doubt $\varepsilon \in (0,1)$, any "correct" posterior $\rho = \mathbb{P}\left[\theta = 1 \mid s^{(1)}, s^{(2)}\right] \in (0, 0.5)$, we can find a prior, a signal structure $F_i^{(\theta)}$, $i \in \{1,2\}$, $\theta \in \{0,1\}$, and the realization of the signals $s^{(1)}$ and $s^{(2)}$ such that the ∞-level agents with $f(0) = \ldots = f(k-1) = 0$ and $f(k) = \varepsilon$ permanently disagree with each other.

size of the doubt

Theorem

For any level $k \geq 1$, any positive amount of doubt $\varepsilon \in (0, 1)$, any "correct" posterior $\rho = \mathbb{P}\left[\theta = 1 \mid s^{(1)}, s^{(2)}\right] \in (0, 0.5)$, we can find a prior, a signal structure $F_i^{(\theta)}$, $i \in \{1, 2\}$, $\theta \in \{0, 1\}$, and the realization of the signals $s^{(1)}$ and $s^{(2)}$ such that the ∞-level agents with $f(0) = \ldots = f(k-1) = 0$ and $f(k) = \varepsilon$ permanently disagree with each other.

size of the doubt

highest level of rationality this doubt is placed upon

Theorem

For any level $k \geq 1$, any positive amount of doubt $\varepsilon \in (0,1)$, any "correct" posterior $\rho = \mathbb{P}\left[\theta = 1 \mid s^{(1)}, s^{(2)}\right] \in (0, 0.5)$, we can find a prior, a signal structure $F_i^{(\theta)}$, $i \in \{1,2\}$, $\theta \in \{0,1\}$, and the realization of the signals $s^{(1)}$ and $s^{(2)}$ such that the ∞-level agents with $f(0) = \ldots = f(k-1) = 0$ and $f(k) = \varepsilon$ permanently disagree with each other.

size of the doubt

highest level of rationality this doubt is placed upon

"correct" posterior belief $\rho \in (0, 0.5)$: $0.5 - \rho =$ ex post disagreement

Theorem

For any level $k \geq 1$, any positive amount of doubt $\varepsilon \in (0,1)$, any "correct" posterior $\rho = \mathbb{P}\left[\theta = 1 \mid s^{(1)}, s^{(2)}\right] \in (0, 0.5)$, we can find a prior, a signal structure $F_i^{(\theta)}$, $i \in \{1,2\}$, $\theta \in \{0,1\}$, and the realization of the signals $s^{(1)}$ and $s^{(2)}$ such that the ∞-level agents with $f(0) = \ldots = f(k-1) = 0$ and $f(k) = \varepsilon$ permanently disagree with each other.

size of the doubt

highest level of rationality this doubt is placed upon

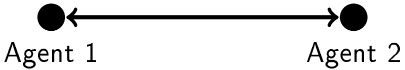
"correct" posterior belief $\rho \in (0, 0.5)$: $0.5 - \rho =$ ex post disagreement

$\Downarrow$

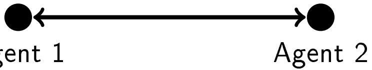
ex ante disagreement needed to guarantee

this ex post disagreemen

Example 1



Example 1

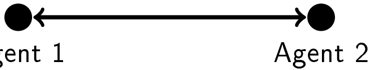


Signals: 0 1

Common prior: ρ. Signals are binary and iid:

$$\mathbb{P}[s = \theta \mid \theta] = q \in (0.5, 1), \quad \mathbb{P}[s = 1 - \theta \mid \theta] = 1 - q$$

Example 1



	Agent 1	Agent 2
Signals:	0	1
Period 1:	0	1

Assume $\rho + q > 1 \Rightarrow$ always optimal to guess the signal in period 1: $\mathbb{P}\left[\theta = s \mid s\right] > 0.5$

Example 1

	Agent 1	Agent 2
Signals:	0	1
Period 1:	0	1
Period 2:	0	
	⋮	
	0	

Prior is biased towards 0 $\Rightarrow$ agent 1 always reports 0

Example 1

	Agent 1	Agent 2
Signals:	0	1
Period 1:	0	1
Period 2:	0	?
	⋮	
	0	

Agent 2:

$f(0)$ → agent 1 is of level 0 $\Rightarrow$ best guess is 1

$1 - f(0)$ → agent 1 is of level ≥ 1 $\Rightarrow$ best guess is 0

Example 1

	Agent 1	Agent 2
Signals:	0	1
Period 1:	0	1
Period 2:	0	? $\leftarrow$ depends on $f(0)$
	$\vdots$	
	0	

Agent 2:

$f(0)$ $\rightarrow$ agent 1 is of level 0 $\Rightarrow$ best guess is 1

$1 - f(0)$ $\rightarrow$ agent 1 is of level $\geq 1 \Rightarrow$ best guess is 0

Example 1

	Agent 1	Agent 2	
Signals:	0	1	
Period 1:	0	1	
Period 2:	0	?	$\leftarrow$ depends on $f(0)$
	$\vdots$	$\vdots$	
	0	0	$\leftarrow$ agreement is inevitable

$$\mathbb{P}\left[s^{(2)} = 1,\ y_1^{(1)} = \ldots = y_{t-1}^{(1)} = 0 \mid \text{agent 1 is of level 0}\right] \underset{t\to\infty}{\longrightarrow} 0$$

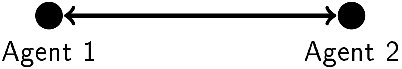
$\Rightarrow$ eventually, agent 2 learns that agent 1 is of level ≥ 1

Example 1

	Agent 1	Agent 2	
Signals:	0	1	
Period 1:	0	1	
Period 2:	0	?	$\leftarrow$ depends on $f(0)$
	$\vdots$	$\vdots$	
	0	0	$\leftarrow$ agreement is inevitable

1. Absence of c.k. of rationality does not necessarily lead to disagreement
2. Agents might learn each other's level of rationality

Example 2



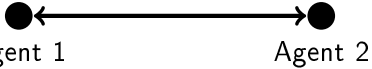
Signals: $\{0, 0\}$ $\{1, 1\}$

Common prior: ρ. Signals are binary and iid:

$$\mathbb{P}\left[s = \theta \mid \theta\right] = q \in (0.5, 1), \quad \mathbb{P}\left[s = 1 - \theta \mid \theta\right] = 1 - q$$

$\Rightarrow$ each agent receives $s = s_1 + s_2 \sim B(2, q)$

Example 2



	Agent 1	Agent 2	
Signals:	$\{0, 0\}$	$\{1, 1\}$	
Period 1:	0	1	$\longleftarrow \rho + q > 1$

Example 2

	Agent 1	Agent 2	
Signals:	$\{0, 0\}$	$\{1, 1\}$	
Period 1:	0	1	$\longleftarrow \rho + q > 1$
Period 2:	0		
Period 3:	0		
	$\vdots$		
	0		

Prior is biased towards 0 $\Rightarrow$ agent 1 always reports 0

Example 2

	Agent 1	Agent 2	
Signals:	$\{0, 0\}$	$\{1, 1\}$	
Period 1:	0	1	$\longleftarrow \rho + q > 1$
Period 2:	0		
Period 3:	0		
	$\vdots$		
	0		

Assume $f(0) = 0$

Example 1 $\Rightarrow$ f(0) only impends learning, though not causes disagreement per se.
Example 2 shows disagreement $\Rightarrow$ setting $f(0) = 0$ makes our argument stronger

Example 2

	Agent 1	Agent 2	
Signals:	$\{0, 0\}$	$\{1, 1\}$	
Period 1:	0	1	$\longleftarrow \rho + q > 1$
Period 2:	0	1	
Period 3:	0		
	$\vdots$		
	0		

Assume $f(0) = 0$

Agent 2: agent 1 is of level $\geq 1 \Rightarrow s_1^{(1)} + s_2^{(1)} \leq 1$

$$\mathbb{P}\left[\theta = 1 \mid s_1^{(2)} = s_2^{(2)} = 1,\ s_1^{(1)} + s_2^{(1)} \leq 1\right] > \frac{1}{2}$$

Example 2

	Agent 1	Agent 2	
Signals:	$\{0,0\}$	$\{1,1\}$	
Period 1:	0	1	$\longleftarrow \rho + q > 1$
Period 2:	0	1	
Period 3:	0	1	$\longleftarrow$ if $f(1) > \frac{1}{2\left(\frac{\rho+q-1}{(1-2\rho)(1-q)q}+1\right)}$
	$\vdots$		
	0		

Assume $f(0) = 0$

$f(1)$ → agent 1 is of level 1 $\Rightarrow s_1^{(1)} + s_2^{(1)} \leq 1 \Rightarrow$ best guess is 1

$1 - f(1)$ → agent 1 is of level $\geq 2 \Rightarrow s_1^{(1)} = s_2^{(1)} = 0 \Rightarrow$ best guess is 0

Example 2

	Agent 1	Agent 2	
Signals:	$\{0,0\}$	$\{1,1\}$	
Period 1:	0	1	$\longleftarrow \rho + q > 1$
Period 2:	0	1	
Period 3:	0	1	permanent disagreement if $f(1) > \frac{1}{2\left(\frac{\rho+q-1}{(1-2\rho)(1-q)q}+1\right)}$
	$\vdots$	$\vdots$	
	0	1	

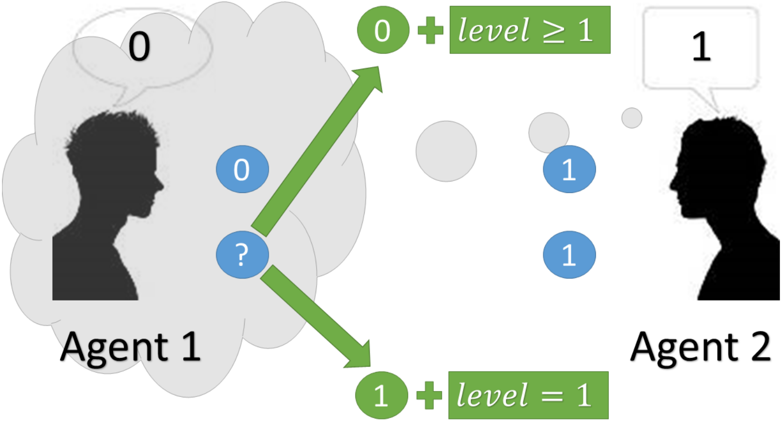
Assume $f(0) = 0$

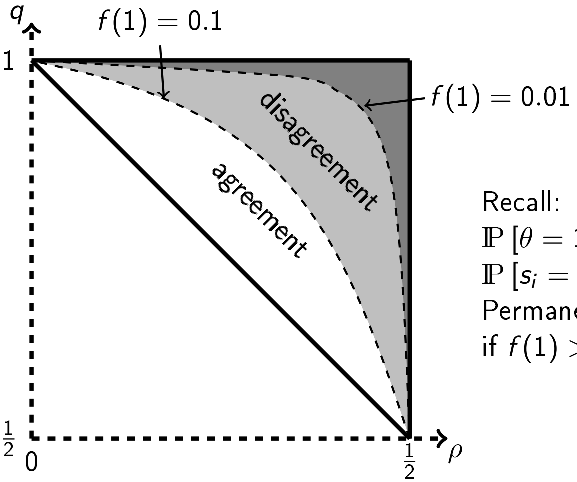
Agent 2 knows:

- level 1 agent 1 stops updating after period 1
- level ≥ 2 agent 1 stops updating after period 2
- $\Rightarrow$ agent 1 stops updating after period 2

$\Rightarrow$ agent 2 stops updating after period 3

Example 2





Recall:

$\mathbb{P}[\theta = 1] = \rho$

$\mathbb{P}[s_i = \theta \mid \theta] = q$

Permanent disagreement

if $f(1) > \frac{1}{2\left(\frac{\rho+q-1}{(1-2\rho)(1-q)q}+1\right)}$

The size of the doubt $f(1)$ can be arbitrary small and the correct posterior ρ can be arbitrary close to 0: for any $f(1) \in (0,1)$ and any $\rho \in (0, 0.5)$ we can find $q \in (0.5, 1)$ (characterizes ex ante disagreement) such that the agents permanently disagree with each other for some signals realizations.

Conclusion

Paradox

Aumann (1976): impossibility to agree to disagree

Our result

Model with almost common knowledge of rationality where disagreement is possible